UNTERSUCHUNG DER SPEKTRALSTRUKTUR VON /f/-LAUTEN UND MÖGLICHKEITEN IHRER NUTZUNG IN DER FORENSISCHEN PHONETIK

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu
BONN

vorgelegt von

Dr. rer. nat. GERHARD VAN DER GIET
aus
SCHAEPHUYSEN

BONN 1986

UNTERSUCHUNG DER SPEKTRALSTRUKTUR VON /f/-LAUTEN UND MÖGLICHKEITEN IHRER NUTZUNG IN DER FORENSISCHEN PHONETIK

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu
BONN

vorgelegt von

Dr. rer. nat. GERHARD VAN DER GIET
aus
SCHAEPHUYSEN

BONN 1986

1. Referent: Prof. Dr. W. Lenders

2. Referent: Prof. Dr. H. P. Kelz

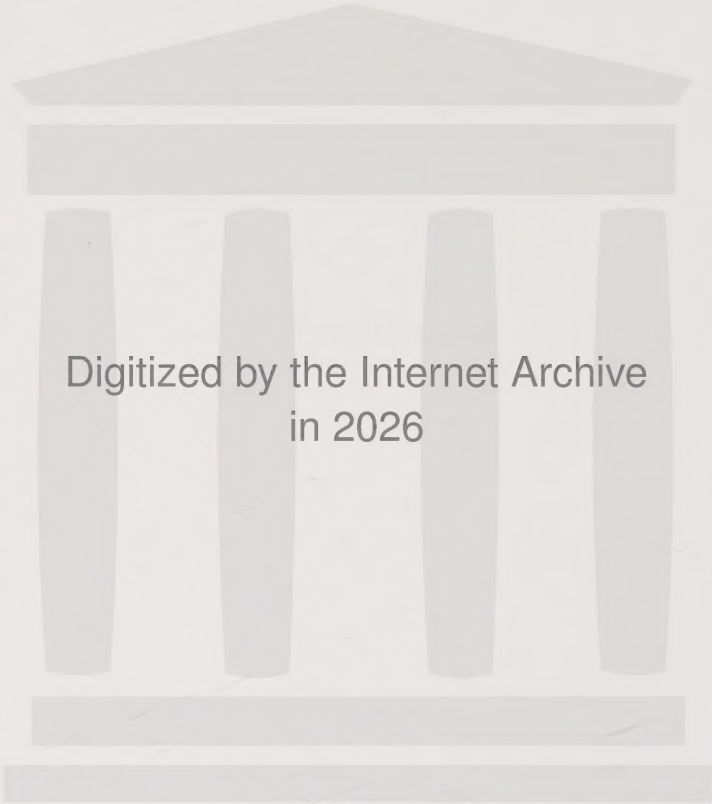
Tag der mündlichen Prüfung: 11. Dezember 1985

Die Arbeit wurde angefertigt am

BUNDESKRIMINALAMT
Wiesbaden

Druck: Rheinische Friedrich-Wilhelms-Universität Bonn

Meinen Eltern


Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41546505_0021

INHALTSVERZEICHNIS

Seite

Zusammenfassung 8

1. Maschinelle Sprachverarbeitung und Forensische Phonetik 9

1.1 Schwerpunkte maschineller Sprachverarbeitung 9
1.2 Aufgaben der forensischen Phonetik 14
1.3 Nutzung von Sprachlautspektren in der forensischen Phonetik 18

2. Einfluß des Übertragungskanals auf das Sprachsignal .. 23
und Möglichkeiten der Kompensation 23

2.1 Die Übertragungsstrecke 23
2.2 Möglichkeiten der Kanalkompensation 27
2.3 Kanalvermessung mithilfe von Sprachlautspektren .. 29
2.4 Zum Gegenstand der vorliegenden Untersuchung 30

3. Spektrale Struktur der /f/-Laute 33

3.1 Akustische Überlegungen 33
3.2 Messungen zur Spektralstruktur 37
3.2.1 /f/-Spektren in lautlichem Kontext 37
3.2.2 /f/-Spektren in Isolation 42
3.2.3 Vergleichende Gegenüberstellung verschiedener Messungen 47

4. Berechnung der /f/-Spektren 49

4.1 Mathematische Grundlagen 49
4.1.1 Anmerkungen zur Fouriertransformation 49
4.1.2 Anwendung der Fouriertransformation auf stochastische Signale 53
4.1.3 Die Fast-Fourier-Transformation 57

Seite

4.2 Vorgehensweise und exemplarische Ergebnisse 59

4.2.1 Einzelspektren 59
4.2.2 Mittelwertspektren 64
4.2.3 Normierung der Spektren 66
4.2.4 Art der Mittelung 66
4.2.5 Auswahl des Datenmaterials und der Berechnungsmethode 67

5. Ergebnisse 71

5.1 Dauer von /f/-Lauten 71
5.2 Spektren der /f/-Realisationen 77
5.2.1 Gegenüberstellung der Spektren 77
5.2.2 Typ1: Mittelung über Kontexte 79
5.2.3 Typ2: Mittelung über Realisationen 81
5.2.4 Typ3: Mittelung über Realisationen und Kontexte 83
5.2.5 Typ4: Mittelung über Sprecher und Realisationen 85
5.2.6 Typ5: Mittelung über Sprecher, Realisationen und Kontexte 87

5.3 Einfluß des lautlichen Kontextes 89

5.3.1 Segmentation der Lautumgebungen 89
5.3.2 Kontext 1: /f/ in /Vater/ 90
5.3.3 Kontext 2 und 3: /f/ in /schläft viel/ 92
5.3.4 Kontext 4: /f/ in /fach/ 92
5.3.5 Kontext 5: /f/ in /vordem/ 94
5.3.6 Kontext 6: /f/ in /Fernseher/ 94
5.3.7 Kontext 7: /f/ in /fällt/ 94
5.3.8 Kontext 8: /f/ in /Familie/ 96
5.3.9 Kontext 9: /f/ in /auf/ 96

Seite

5.4 Streubreiten der /f/-Spektren 98
5.4.1 Berechnung der Streubreiten 98
5.4.2 Mittlere Streubreiten für unterschiedliche Mengen von /f/-Realisationen 104

6. Kriminaltechnische Anwendungen 115

6.1 Überblick 115
6.2 Datensammlung 117
6.3 Messungen an simulierten Übertragungskanälen mit Filtern 120
6.4 Kontextunabhängige Messung an echten Telefonkanälen 122
6.4.1 Ortsleitungen mit gewöhnlichen Fernsprechapparaten 124
6.4.2 Ortsleitungen mit öffentlichen Münzfernsprechern 127
6.4.3 Leitungen aus dem Nahbereich 127
6.4.4 Leitungen aus der Fernebene 127

6.5 Kontextabhängige Messung an echten Telefonkanälen 132
6.6 Identifizierung von Fernsprechapparaten 135
6.7 Ent- und Verzerrung von Sprachaufzeichnungen für den Stimmenvergleich 137

7. Schlußfolgerungen und Ausblick 139

8. Literaturdokumentation 140

Lebenslauf 161

ZUSAMMENFASSUNG

In dieser Arbeit wird ein neues Verfahren zur nachträglichen Messung der Frequenzeigenschaften einer Übertragungsstrecke allein aus dem Sprachsignal vorgestellt. Es werden detaillierte signalphonetische Erkenntnisse im Hinblick auf eine Anwendung im Bereich der kriminaltechnischen Sprachverarbeitung über die Frequenzeigenschaften bestimmter Sprachlaute gesammelt und Untersuchungen durchgeführt, inwieweit diese Laute zur nachträglichen Messung herangezogen werden können. Für diesen Zweck besonders geeignet und damit zentraler Gegenstand dieser Untersuchung ist der Frikativ /f/, der aufgrund seines labio-dentalen Artikulationsortes nur wenig durch das Ansatzrohr bedingte und daher auch sprecherspezifische Information aufnehmen kann. Aus den Abweichungen der Spektren der /f/-Realisierungen in telefonübertragener Sprache von den entsprechenden für /f/-Realisierungen charakteristischen vorgegebenen Spektren läßt sich die Übertragungsfunktion unter gewissen Voraussetzungen rekonstruieren. Die akustische Struktur der /f/-Laute wird eingehend untersucht und das Datenmaterial für die Vermessung selbst sowie für die Bewertung ihrer Zuverlässigkeit bereitgestellt. In diesem Zusammenhang wird insbesondere auf die Streuungen der Spektren unter verschiedenen Randbedingungen sowie auf Kontexteinflüsse eingegangen. Anhand von praxisgerechten Beispielen wird gezeigt, daß das Verfahren aller Voraussicht nach für den Einsatz im Bereich der forensischen Sprachverarbeitung geeignet ist. Der Arbeit ist eine umfangreiche Literaturdokumentationbeigefügt, um dem Leser eine Vielfalt an relevanten Informationen zum genannten Thema anzubieten.

1. MASCHINELLE SPRACHVERARBEITUNG UND FORENSISCHE PHONETIK

1.1 Schwerpunkte maschineller Sprachverarbeitung

In den letzten Jahren hat die Verarbeitung von Sprache durch Maschinen zunehmend an Bedeutung gewonnen. Grosse Anstrengungen bei der Entwicklung von Spracherkennungsgeräten und Systemen zur Produktion synthetischer Sprache bei Forschungsinstituten und in der Industrie haben zu ersten einsatzfähigen Systemen geführt. Angegebene - allerdings höchstens zum Vergleich untereinander geeignete - Erkennungsleistungen von über 95 % haben einen Markt geweckt, der nach Schätzungen gegen Ende des Jahrzehntes die Milliarden Dollar Grenze überschreiten soll (DIEBOLD 1981). Synthetische Sprache ist zwar nach wie vor anhand ihres von normaler Sprache abweichenden Klanges als solche zu erkennen; ihre Verständlichkeit ist aber ausreichend gut, so daß ihr Einsatz zur Informationsausgabe unter den verschiedensten Anwendungsbedingungen erprobt wird.

Sowohl im zivilen (z.B.: WHITE et. al. 1976, VAN DER GIET 1977, VAN DER GIET 1978, GLAVE et. al. 1978) wie auch im militärischen Bereich (z.B.: BEEK 1977, NADC 1980, VAN DER GIET 1979, AGARD 1982, NATO 1983) besteht außerordentliches Interesse an der Nutzung der Sprache als Kommunikationsmedium für den Dialog mit Maschinen (Spracherkennung und -synthese) und als ein für viele Zwecke sinnvolles Identifikationsmedium (Sprechererkennung) (z.B.: UNGEHEUER 1965, UNGEHEUER 1966, ATAL 1972, SAMBUR 1975, VAN DER GIET et.al. 1981, MUTSCHLER 1983). Die Flugzeugindustrie ist ein Beispiel für einen potentiellen Anwender mit höchstem Interesse an Sprachverarbeitungstechnologie, da deren Einsatz z.B. im Cockpit zu ganz erheblichen Verbesserungen des Mensch-Maschine-Systems "Flugzeug" führen würde. Es haben sich jedoch aus wissenschaftlicher wie auch aus der Sicht des Anwenders eine Reihe von ungelösten Problemen ergeben, die den praktischen Einsatz der Sprachverarbeitung nachhaltig behindern. Dazu gehört bei der Spracherkennung ihre Sprecherabhängigkeit, die

in der Regel jedes Mal vor Aufnahme einer Nutzanwendung die individuelle Adaption des Systems auf den einzelnen Sprecher erfordert.

Die Entwicklung eines sprecherunabhängigen Spracherkennungssystems erfordert ausreichende Kenntnisse über die Unterschiede von Sprecher zu Sprecher bei der Realisierung von Sprache. Sind die sprecherspezifischen Unterschiede einmal bekannt, können sie in der Spracherkennung zur sogenannten "Sprechernormalisierung", bei der Sprechererkennung dagegen als Mittel zur Identifizierung benutzt werden. Hier verdeutlicht sich die enge Verzahnung beider Bereiche, in denen methodisch gleichartig vorgegangen wird. Zur Zeit wird gerade an dieser Problematik weltweit gearbeitet, da zahlreiche Institutionen mehr oder weniger leistungsfähige Spracherkennungsgeräte entwickelt haben, der mögliche Nutzen durch die Sprecherabhängigkeit aber zu sehr eingeschränkt wird.

Ein ebenfalls nicht befriedigend gelöstes Problem ist die Erkennungssicherheit bei telefonübertragener Sprache. Es gibt auch hier Hersteller, die für ihre Systeme volle "Funktionsfähigkeit" beanspruchen. Hierunter wird in der Regel lediglich die Einhaltung hoher Erkennungsraten verstanden. Eine Aussage über die praktische Nutzbarkeit bezogen auf eine konkrete Anwendung ist damit nicht zwangsläufig verbunden; auf diese Problematik kann jedoch hier nicht näher eingegangen werden. Grundsätzlich ist Angaben über Erkennungsraten, die in aller Regel auf Labormessungen beruhen, mit äußerster Vorsicht zu begegnen. Realistische Bedingungen, wie sie normalerweise beim Telefonieren auftreten (unbekannte Leitungseigenschaften, unterschiedliches Halten des Telefonhörers, Hintergrundgeräusche, etc.), sind nur bedingt simulierbar. Das bedeutendste Problem ist dabei nicht einmal die verringerte Erkennungsleistung - auf diese könnte man sich einstellen - sondern ihre Unvorhersehbarkeit, da neben der bereits genannten Qualität bzw. den Eigenschaften der aktuellen Übertragungsstrecke auch das individuelle Kommunikationsverhalten am Telefon in der betreffenden

Kommunikationssituation und eine Reihe von weiteren Effekten die Erkennungsleistung unmittelbar beeinflussen.

Das zentrale Problem bei der Sprach- und Sprechererkennung ist die "Variabilität" (LEA 1982). Solange nicht bekannt ist, welche Faktoren die gesprochene Sprache in welcher Weise beeinflussen, kann nur versucht werden, durch Festlegung der Randbedingungen soweit möglich einen bekannten und kontrollierbaren Rahmen zu schaffen. Im Gegensatz zu später diskutierten kriminaltechnischen Anwendungen kann nämlich (z.B. durch entsprechende Anweisungen) weitgehend dafür gesorgt werden, daß eine hohe Sprechdisziplin gewährleistet wird, daß Hintergrundgeräusche auf ein Minimum reduziert werden, usw. Außerdem trägt eine durch Übung mit dem System gewonnene Adaption des Sprechers an das System zur Erkennungssicherheit bei. Die Sprecher sind kooperativ, da sie zwangsläufig Interesse an einem erfolgreichen Erkennungsprozeß haben.

Die Kooperativität der Sprecher manifestiert einen entscheidenden Unterschied zwischen dieser manchmal auch "kommerziell" oder "industriell" genannten Art der Sprachverarbeitung und entsprechenden Aktivitäten im kriminaltechnischen Bereich, da dort im allgemeinen von Nichtkooperativität (z.B. Stimmverstellung oder Immitation) ausgegangen werden muß. Die Termini "kommerziell" bzw. "industriell" beinhalten meist auch gleichzeitig "automatisch". Die Entscheidung, um welches Wort, um welchen Satz oder um welchen Sprecher es sich handelt, trifft die Maschine ohne jeden Eingriff des Menschen in den Erkennungsprozeß. Die automatische Erkennung, die meist sogar in Realzeit gefordert werden muß, ist eine Voraussetzung für die meisten praktischen Anwendungen. Auch hierin liegt eine Abgrenzung gegenüber der Sprechererkennung in der Kriminaltechnik, bei der automatische Entscheidungen - wie an späterer Stelle ausführlicher dargelegt wird - zur Zeit nicht gerichtsverwertbar realisiert werden können und bei der Zeitfragen der hier betrachteten Größenordnung keine Rolle spielen.

Trotz der umfangreichen Forschungs- und Entwicklungsarbeiten und verhältnismäßig günstigen Ausgangsbedingungen (kontrollierbare Kommunikationssituation) hat die tatsächliche Performanz selbst der kommerziellen Sprachverarbeitung wegen der genannten Probleme mit den eingangs geschilderten Erwartungen nur zum Teil Schritt halten können:

> "Will all these expectations come true? I am not the one to judge but there is certainly room for some scepticism or at least caution. The rate of increase of the world market has not progressed at the expected rate. So - the surf on the tidal wave of expectations may end in a brake when we are confronted with the reefs of the knowledge barrier. I am referring to our still meager inside in speech as a language code. We need a fifth generation of speech scientists rather than a fifth generation of computers." (FANT 1983)

Wie bei allen automatischen Erkennungssystemen werden auch bei der Erkennung von Sprache oder Sprecher aus dem aktuellen Sprachsignal Muster abgeleitet und mit früher (in der Lernphase) gewonnenen und gespeicherten verglichen. Während die zweite Stufe des Erkennungsprozesses (back end) im allgemeinen mit den üblichen Methoden der Mustererkennung durchgeführt wird, ist bei der Ableitung der Muster selbst (front end) eine besonders große Breite von anwendbaren Techniken und Verfahren zu beobachten. Sie reichen von mehr nachrichtentechnischen (z.B. Filterbankwerte, LPC) bis hin zu phonetischen Beschreibungskategorien (z.B. im DAWID-Spracherkennungssystem, siehe VAN DER GIET 1978).

Das Beispiel der automatischen Spracherkennung zeigt, daß die Art der Problemlösung - so einleuchtend sie zunächst scheint - zwangsläufig auf eine Wissensbarriere stoßen mußte:
Lediglich aus dem Sprachsignal mit den gegebenen technischen Möglichkeiten abstrakte Muster ableiten und als Referenz für

die Erkennung bereitstellen liefert zwar eine gewisse Performanz, doch:

> "We learn very little about speech this way and we are generally not aware of why the matching incidentally fails. The other approach needed for large vocabularies and connected speech is phonetically oriented in the sense that it is based on recognition of minimal units that can range from distinctive features, phonemes, diphones, syllables, and words and which require some kind of segmentation. We now approach the general problem of speech research in quest of the speech code and the relation between message units and their phonetic realization with all the variability induced by contextual factors including language, dialect, speaker specific situational and stylistic variations." (FANT 1983)

Die fehlende Kenntnis über menschliche Perzeption setzt der Entwicklung leistungsfähiger Spracherkennungssysteme enge Grenzen. Einer der Wege, die Fant's "Fifth Generation Scientists" beschreiten könnten, dürfte deshalb zur Umgehung dieses Problems die Zusammenführung des Menschen mit seiner Kompetenz und technischen Einrichtungen zu einem Gesamtsystem sein, das in seiner Leistungsfähigkeit und Einsatzbreite neue Möglichkeiten zur Forschung und Anwendung eröffnen kann. Ein solcher Weg wird in dieser Arbeit für den Einsatz von Sprachverarbeitungstechnologie im Bereich der Kriminaltechnik beschritten: Durch Voruntersuchungen wird umfangreiches Hintergrundmaterial ermittelt und dem Experten in Verbindung mit einem technischen System zur Verfügung gestellt. Die Zusammenführung dieser drei Komponenten, der Experte, das technische System sowie die Datenbank mit spezifischen Informationen, könnte einen Ansatz darstellen, der sich gleichermaßen zur Lösung von Forschungs- und Anwendungsfragen eignet, die in einem gewissen Sinne "offen" sind, d.h. bei denen sich durch neue oder neue Kombinationen

von Randbedingungen immer wieder neue Fragestellungen ergeben. In einem solchen Konzept ist auch die hier vorgestellte Methode zur Bestimmung von Kanaleigenschaften zu sehen.

Zum Verständnis der Problematik dieser Arbeit wird in den folgenden Abschnitten des ersten Kapitels zunächst eine "Forensische Phonetik" durch Beschreibung ihrer Aufgaben umrissen. Daraus ergibt sich als Teilproblem die in dieser Arbeit behandelte Aufgabenstellung, die im letzten Abschnitt dieses Kapitels detaillierter dargestellt wird.

1.2 Aufgaben der forensischen Phonetik

Vor dem Hintergrund der geschilderten Möglichkeiten moderner technischer Sprachanalyse und -verabeitung ist auch in der Kriminaltechnik der Versuch gemacht worden, durch entsprechende Verfahren zur Aufklärung von Straftaten beizutragen, bei denen die Sprache als Kommunikationsmedium, z.B. zwischen Täter und Opfer, eingesetzt wurde. Die Fragen, deren Beantwortung sich Polizei und Gerichte durch Einsatz von Sprachverarbeitungstechnologie erhoffen, sind zwar vielfältig, lassen sich jedoch für viele Kriminalfälle aufgrund des in gewisser Weise ähnlichen Ablaufes der meisten Fälle in typische Kategorien zusammen fassen: wenn eine Straftat unter Zuhilfenahme sprachlicher Kommunikation begangen wird (z.B. Erpressungsanruf nach einer Entführung), ergibt sich für die Polizei unmittelbar der Wunsch nach Angaben über den Sprecher. Seiner Sprache können tatsächlich oft viele Informationen entnommen werden, z.B. Herkunft (Dialekt), Alter, Geschlecht, Bildungsniveau, Sprechsituation (Hintergrundanalyse), psychologische Situation (Stress), usw. Später finden sich Verdächtige, die in der Regel ihre Identität mit dem Sprecher des Tatanrufes bestreiten und deshalb bereitwillig Proben für einen Stimmenvergleich abgeben. Da nun der Eindruck des ungeschulten Hörers (z.B. des ermittelnden Kriminalbeamten) weder für die Polizei noch für

die Gerichte ausreichen kann, muß diese Aufgabe an den Experten weitergegeben werden. Die von Gutachtern angefertigten Expertisen sind nun aber zwangsläufig bis zu einem gewissen Grad subjektiv. Gerade in den letzten Jahren haben nun verstärkt Bemühungen eingesetzt, Strafverfahren zu "objektivieren", d.h. soweit möglich menschliche Unzulänglichkeiten (meist in Zeugenaussagen, aber auch in Gutachten) dadurch für die Prozesse unbedeutend zu machen, daß zunehmend qualifizierte Sachbeweise möglich werden. In diesem Rahmen wurde gerade bei dem für die Kriminaltechnik - von einzelnen Versuchen in der Vergangenheit abgesehen - jungen Gebiet der forensischen Sprachverarbeitung durch ein Höchstmaß an eingesetzter Technik versucht, Systeme zu schaffen, die nicht nur komplexe Meßinstrumente darstellen, sondern wie bei der kommerziellen Sprechererkennung und auf diese Erfahrungen aufbauend Entscheidungskompetenz übernehmen sollten. Bis heute sind sämtliche dieser Versuche ohne Erfolg geblieben, da die Randbedingungen im kriminaltechnischen Bereich derart vielfältig und diffizil sind, daß das derzeitige Wissen über die sprachliche Kommunikation eine solche Entwicklung nicht zuläßt.

In automatischen Sprechererkennungssystemen wird für die Entscheidung die Distanz zwischen dem Merkmalsvektor, der den Sprecher der aktuellen Stimmprobe beschreibt, mit gespeicherten verglichen. Im kommerziellen Bereich läßt sich immer eine in einem gewissen Sinn "abgeschlossene Welt" festlegen: so sind z.B. die Personen, die zur Erkennung infrage kommen, bekannt; sie geben die Stimmprobe unter im wesentlichen bekannten Bedingungen ab, usw. Typisch wäre die Bedienung eines automatischen Zugangskontrollsystems in einer akustisch geeigneten Kabine. Die Entscheidung beruht auf einer Interpretation (Vergleich mit Schwellen) der gemessenen Distanz, die dadurch möglich wird, daß die in dieser "Welt" auftretenden Distanzen in Form von Statistiken bekannt sind. Im forensischen Bereich gibt es diese Abgeschlossenheit nicht. Die nahezu unbegrenzte Zahl unkontrollierbarer Einflüsse wie die unbekannte Sprechsituation, Stress, Störungen durch Hintergrundgeräusche,

Telefonübertragung, usw. können nicht in endlich vielen Statistiken niedergelegt werden. Verstärkt wird dieses Problem zusätzlich durch die fehlende Kooperativität der Sprecher, die dazu führt, daß grundsätzlich mit allen Maßnahmen gerechnet werden muß, von denen sich ein Sprecher eine Erschwerung seiner Identifikation erhoffen kann. Stimmverstellungen und -immitationen oder Veränderungen, z.B. durch mechanische Hilfsmittel, werden bei Kriminalfällen häufig beobachtet.

Ein kriminaltechnisches Erkennungssystem müßte die Vielfalt dieser Möglichkeiten und ihre Auswirkungen auf die zur Identifikation benutzten sprecherspezifischen Merkmale "kennen", und ihre Eigenschaften in Hintergrundstatistiken bereitstellen. Zu ihrer Ermittlung müßte das unmögliche Vorhaben realisiert werden, sämtliche Kommunikationsituationen und technischen Randbedingungen in einem Labor zu simulieren und entsprechende Analysen durchzuführen. Die Distanz zwischen Labor und realer Welt läßt daher auch in fernerer Zukunft die Entwicklung eines technischen Systems mit gerichtsverwertbarer Entscheidungskompetenz für oder gegen die Identität von Sprechern als unrealistisch erscheinen. Es bleibt somit nur die Analyse des Experten, der zwar das technische System als Meßinstrument zur Entscheidungsfindung benutzen kann, im wesentlichen den Stimmenvergleich aber auf der Basis einer auditiv-phonetischen Analyse in Verbindung mit seinem Expertenwissen durchführt.

Der Ablauf eines solchen Stimmenvergleiches unterscheidet sich erheblich von denen technischer Systeme für die Erkennung kooperativer Sprecher. Begriffe wie Sprecheridentifikation (Ermittlung der Identität aus einer Menge von Sprechern, ein Mehrklassenproblem) und Sprecherverifikation (Bestätigung einer Identität, ein Zweiklassenproblem) sind hier nicht mehr sinnvoll, da immer nur Paarvergleiche vorgenommen und die Ergebnisse gegeneinander abgewogen werden. Um diesen Besonderheiten Rechnung zu tragen, wird im folgenden für die durch den Experten vorgenommene Sprechererkennung ausschließlich der Begriff

"Stimmenvergleich" verwendet.

In Anbetracht der Möglichkeiten moderner Meßtechnik sollte - wie bereits angedeutet - bei der auditiv-phonetischen Analyse auf apparative Unterstützung nicht verzichtet werden. Da bereits zahlreiche Versuche beschrieben worden sind, insbesondere auch Computer zur kriminaltechnischen Sprecherer-kennung, Sprachverbesserung, usw. nutzbar zu machen, (z.B.: BUNDESKRIMINALAMT 1980, BUNDESKRIMINALAMT 1981, VAN DER GIET et.al. 1980, VAN DER GIET et.al. 1981, LEWIS 1980, BLAIN 1980) wird heute oft von forensischer Sprachverarbeitung gesprochen, wobei der Begriff "forensisch" seiner ursprünglichen Bedeutung nach auf die gerichtliche Verwertbarkeit hinweist, meist aber - so auch in dieser Arbeit - zur Bezeichnung des gerichtlichen und kriminaltechnischen Bereiches gebraucht wird.

Der vom Experten, meist einem Phonetiker, durchgeführte auditiv-phonetische Stimmenvergleich, dessen Ablauf im einzel-nen hier nicht wiedergegeben werden kann, umfaßt im Prinzip alle Disziplinen der Phonetik wie auch benachbarter Wissenschaf-ten. Obgleich seit Bestehen des Telefons bzw. der Möglichkeit zur Aufzeichnung von Sprache immer wieder phonetische Begutach-tungen in Kriminalfällen und bei Gericht vorgetragen wurden, haben sich erst in jüngster Zeit namhafte Institute und zahl-reiche Wissenschaftler mit dieser Problematik befaßt und ein entsprechendes Wissen entwickelt, das sich als eigenständiges Teilgebiet der Phonetik etablieren wird. Die Abgrenzung einer forensischen Komponente innerhalb der Phonetik ergibt sich aus der Kenntnis der Anwendbarkeit einzelner Methoden und der Bewer-tung ihrer Ergebnisse. Die Menge dieses Wissens kann als Forensische Phonetik bezeichnet werden. Damit wird der Vielfalt dieses Gebietes besser Rechnung getragen als in früheren Bezeichnungen wie "Kriminalphonetik", die viel zu sehr den anwendungsbezogenen Aspekt betonten.

1.3 Nutzung von Sprachlautspektren in der forensischen Phonetik

Die Spektren einzelner Laute sowie ihre Veränderungen im Sprachkontinuum bestimmen in maßgeblicher Weise den Höreindruck und werden daher nicht nur vom menschlichen Perzeptionsapparat sondern in Anlehnung daran auch von nahezu sämtlichen technischen Sprach- und Sprechererkennungssystemen zur Identifizierung benutzt. Bei der Übertragung von Sprache über Telefonleitungen werden aber gerade Spektren durch die Übertragungscharakteristik sowie durch Störungen in erheblicher Weise verändert. Die Veränderung von Sprache durch unbekannte Leitungseinflüsse bei der Übertragung ist nicht nur für den Einsatz von Sprachverarbeitungssystemen ein Problem, sondern gerade auch für die auditiv-phonetische Analyse in der Kriminalistik, da die Tataufnahmen fast immer (eine Ausnahme wäre z.B. die direkte "Belauschung" mit Mikrofonen) über Telefon, manchmal auch über Funk, übertragen werden. Grundsätzlich bieten sich zwei Wege zur Lösung dieses Problems: entweder wird versucht, gegen Telefonübertragung resistente Merkmale aus der Stimme abzuleiten (z.B. Zeitmessungen), oder es werden Methoden zur konkreten Ermittlung und Eliminierung der Übertragungseinflüsse entwickelt. Welcher Vorgehensweise der Vorzug gegeben wird, hängt nicht nur von der erwarteten Wirksamkeit der Verfahren ab, sondern zwangsläufig auch von den an der jeweiligen Anwendung orientierten Erfordernissen, die z.B. im kommerziellen Bereich praktisch ausnahmslos eine Realzeit- oder fast Realzeitverarbeitung verlangen. Hier jedoch ist die forensische Phonetik in einer günstigeren Situation als die kommerzielle Sprechererkennung. Der Bearbeiter, z.B. ein Gutachter, hat in der Regel ausreichend Zeit für Untersuchungen. Er kann z.B. auch interaktive Verfahren einsetzen, bei denen der Benutzer in den Verarbeitungsprozeß in Abhängigkeit von seinem Verlauf eingreift.

In dieser Arbeit wird eine interaktive Methode entwickelt, die dem oben an zweiter Stelle genannten Weg entspricht:

Übertragungseigenschaften einer Telefonleitung werden unter bestimmten Voraussetzungen rekonstruiert und anschließend ihr Einfluß eliminiert. Bei diesem Ansatz - im nächsten Kapitel ausführlich beschrieben - wird davon ausgegangen, daß es Sprachlaute gibt, deren Frequenzeigenschaften sich bezüglich verschiedener Sprecher, lautlicher Umgebungen und individueller Realisationen in einem nur engen Bereich verändern. Die nach Übertragung über eine Telefonleitung zu beobachtenden Veränderungen im Spektrum dieser Laute müssen dann zwangsläufig von der Übertragungsstrecke verursacht sein, deren Eigenschaften sich folglich daraus ableiten lassen. Besonders geeignet für ein derartiges Verfahren erscheinen Laute, die aufgrund ihrer Bildungsweise wenig sprecherspezifische Information aufnehmen können. Vokale beispielsweise scheiden allein schon deshalb aus, weil hier die Stimmlippen die Schallquelle bilden und damit das individuell ausgeprägte Ansatzrohr als Resonanzraum diese Laute stark beeinflussen kann. Geeignet dagegen erscheinen Frikative, insbesondere das /f/, das weit vorne artikuliert wird und damit nur wenig von der individuellen Physiologie des Sprechers beeinflußt werden kann. Die Nutzung seines Spektrums zur Ermittlung von Eigenschaften einer Übertragungsstrecke erfordert zunächst eine systematische Untersuchung der spektralen Struktur der /f/-Realisierungen, insbesondere eine Verifizierung der Hypothese, daß ihre Spektren entweder für alle Realisierungen, zumindest aber für bestimmte Realisierungsformen, z.B in bestimmten lautlichen Umgebungen, als hinreichend gleich anzusehen sind. Da dies nur in bestimmtem Umfang zutreffen kann, sind außerdem generell oder für die einzelnen Klassen die Streuungen der Spektren zu ermitteln.

Obgleich das /f/ artikulatorisch eindeutig festgelegt ist, ergeben sich in Abhängigkeit von Kontexten, evtl. auch von Sprechern usw. unterschiedliche Realisierungsklassen, deren Unterscheidung für die vorliegende Arbeit notwendig ist und die im folgenden als /f/-Laute bezeichnet werden. Unter /f/-Realisierungen werden dagegen die tatsächlich vorliegenden Signale als Mitglieder dieser Klassen verstanden.

Bei einer späteren Anwendung erfordert ein solcher phonetisch orientierter Ansatz wegen der sorgfältig durchzuführenden Segmentation ebenfalls einen mit der Signalphänomenologie vertrauten Experten. Im Falle einer lohnenden Anwendbarkeit könnte allerdings anschließend eine Automatisierung durch Entwicklung entsprechender Algorithmen versucht und innerhalb gewisser Grenzen, z.B. bei Beschränkung auf bestimmte Lautklassen und Lautumgebungen, sicher auch erreicht werden.

Die vorgeschlagene Methode läßt sich für eine Reihe von forensischen oder kriminalistischen Fragestellungen einsetzen:

(a) Auditiv-phonetische Analyse

In der auditiv-phonetischen Analyse wird das Tatgespräch - wie erwähnt - mit der Stimme des Verdächtigen unter phonetischen, medizinischen, phoniatrischen, linguistischen, etc., Gesichtspunkten verglichen. Die auditiv-phonetische Analyse ist - dem derzeitigen Stand der beteiligten Wissenschaften entsprechend - im forensischen Bereich die Basis sowohl für die Untersuchung von Einzelstimmen (Dialekt, Alter, etc.) wie auch beim Stimmenvergleich (Prüfung auf Identität). In beiden Fällen, insbesondere aber beim Stimmenvergleich, leidet die Analyse unter den bei der Übertragung entstehenden spektralen Veränderungen. In der Regel (weil für die Polizei am leichtesten zu erstellen) wird das Vergleichsgespräch in einem Büroraum mit recht guter Qualität aufgenommen. Wenn es nun mit der vorgeschlagenen Methode gelingt, nachträglich die Telefonübertragungscharakteristik zu ermitteln, so könnte mit ihrer Hilfe entweder die hochwertige Aufnahme in gleicher Weise verzerrt oder die Telefonsprache entzerrt werden. Letzteres müßte mindestens bei Ferngesprächen auf den Frequenzbereich des Fernleitungskanals beschränkt werden, da oberhalb von 3.4 kHz kein Signal mehr vorliegt und sich sonst eine vereinheitlichende, allenfalls vom Rauschen abhängende Modellierung des Sprachspektrums ergibt.

Die Realisierung von Entzerrung und Verzerrung erfolgt durch entsprechende Filter, die die ermittelte Übertragungsfunktion auf das Sprachsignal anwenden. Hier ist anzumerken, daß in einer Untersuchung (RATHBORN et.al. 1981) die Erkennungsleistung von Menschen bei Vergleich von Telefonsprache mit Telefonsprache bzw. Telefonsprache mit hochwertiger Sprache kein nennenswerter Unterschied festzustellen war und daher die Modifikation der hochwertigen Sprache (z.B. durch Filter) als überflüssig zu betrachten ist. Diese Auffassung wird für den forensischen Bereich vom Autor aus eigener Erfahrung nicht geteilt. Selbst wenn die Ergebnisse aus der genannten Arbeit für die Mehrzahl - zumindest der geschulten - Hörer zutreffen mögen, so beziehen sie sich doch nur auf experimentbezogene Erkennungsraten. Für die forensische Verwendung ist aber nicht nur die Entscheidung selbst relevant sondern auch der - möglicherweise später vor Gericht zu vertretende - subjektive Überzeugungsgrad des Bearbeiters, der zwangsläufig in hohem Maße auf dem in dem betroffenen Stimmenvergleich gewonnenen individuellen Höreindruck beruht.

(b) Entzerrung als Vorstufe zur meßtechnischen Merkmalsextraktion

Schon frühzeitig wurde erkannt (UNGEHEUER 1965, UNGEHEUER 1966), daß viele der wirksamsten Parameter zur Sprechererkennung im spektralen Bereich liegen. Sie verlieren jedoch bei Telefonübertragung der Sprache in ganz besonderem Maße ihre sprecherspezifischen Eigenschaften aufgrund der Veränderungen durch die unbekannte Übertragungscharakteristik. Das hier beschriebene Verfahren könnte der maschinellen (in der Regel durch Computer) Merkmalsextraktion vorgeschaltet werden und somit bewährte Parameter der Sprecheridentifizierung auch bei Telefonsprache nutzbar machen.

(c) Telefonapparatidentifizierung

Eine weitere kriminaltechnische Anwendungsmöglichkeit liegt im Bereich der Telefonapparatidentifizierung. Die Aufgabe ist dabei, im nachhinein zu ermitteln, von welchem einer vorgegebenen Menge von Telefonapparaten aus ein bestimmtes aufgezeichnetes Gespräch geführt wurde. Mit der hier entwickelten Methode kann ein Vergleich der verschiedenen Übertragungswege vorgenommen und so eventuell die Frage beantwortet werden, von welchem Apparat das Gespräch geführt wurde. Dazu muß von jedem der infrage kommenden Fernsprechapparate aus eine Verbindung aufgebaut, die Übertragungscharakteristik ermittelt und mit derjenigen aus dem Tatgespräch verglichen werden. Diejenige mit dem ähnlichsten Übertragungskanal wäre dann als die Tatverbindung anzusehen. Eine zuverläßige Antwort ist allerdings nicht möglich, da sich der identische Verbindungsweg wie beim Tatgespräch wegen des bei jedem Wählvorgang unterschiedlichen Verbindungsaufbaus nicht mehr rekonstruieren läßt. Andererseits bleiben jedoch viele Randbedingungen gleich (z.B. der Fernsprechapparat selbst, die Mikrophonkapsel, der Umgebungsraum, etc.), so daß nach entsprechenden Messungen zumindest brauchbare Fahndungshinweise möglich sind.

2. EINFLUSS DER ÜBERTRAGUNGSSTRECKE AUF DAS SPRACHSIGNAL UND MÖGLICHKEITEN DER KOMPENSATION

2.1 Die Übertragungsstrecke

Die Möglichkeit Sprache auf elektrischem Wege über praktisch beliebige Entfernungen übertragen zu können, macht sie als Identifikationsmedium für viele Anwendungen interessant. Im forensischen Bereich ist aus naheliegenden Gründen praktisch das gesamte anfallende Material über das Telefonnetz übertragen worden, in seltenen Fällen auch über Funk. Dabei unterliegt das Sprachsignal zahlreichen möglichen - und in der Praxis auch meist in unterschiedlich starker Form auftretenden - Störquellen, die für die Zwecke dieser Arbeit ihrer Struktur nach aufzuzeigen sind. Abbildung 2.1 zeigt die verschiedenen signalverändernden Störquellen.

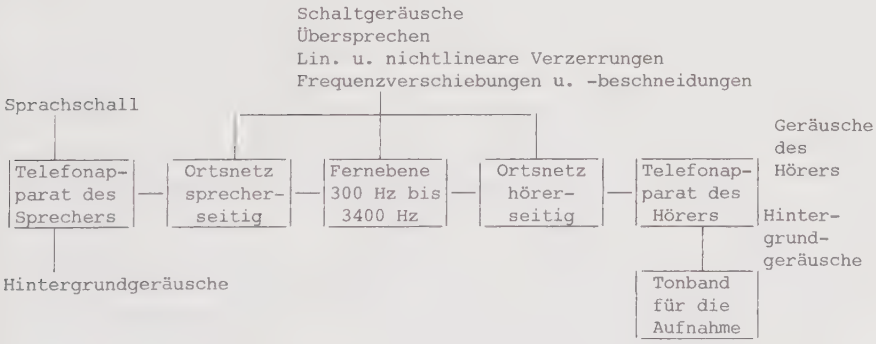


Abb. 2.1: Signalverändernde Störquellen bei der Telefonübertragung

Grundsätzlich muß zwischen auf akustischem Wege auf das Sprachsignal einwirkenden und elektrisch bedingten Veränderungen unterschieden werden. Zu den akustischen Störungen gehören hauptsächlich aus der Umgebung des Sprechers stammende Geräusche. Im Polizeibereich sind dies Straßenlärm, im Hintergrund sprechende Personen, Maschinen, usw. Auch der Sprecher selbst erzeugt u.U. unerwünschte Nebengeräusche wie Scheuern des Hörers an der Kleidung, Atemgeräusche etc., wobei letztere allerdings durchaus als der sprachlichen Kommunikation inherent angesehen werden können. Für die auditiv-phonetische Sprecheridentifizierung sind sie äußerst relevant, da sie oft pathologisch und damit für das Individuum typisch sind. Auch auf der Seite des Hörers können geringfügige akustische Signale über die Sprechkapsel eingekoppelt werden. Im allgemeinen sind diese aber gering, da der Hörer aus Eigeninteresse nicht an zusätzlichen Störungen interessiert ist und Nebengeräusche daher soweit wie möglich unterdrücken wird.

Eine fundamentale Rolle als Störquelle kommt der Kohlekapsel zu, die allerdings - wie erwähnt - nur noch in privaten Telefonapparaten, nicht mehr jedoch in öffentlichen Fernsprechzellen zu finden ist. Die Abbildungen 2.2 (a-c) und 2.3 (a-c) zeigen die Kanaleigenschaften verschiedener Telefonverbindungen. Für die Messungen wurde weißes Rauschen auf die Sprechkapsel des sendenden Apparates gegeben und beim empfangenden Telefon das Spektrum bestimmt. (Bezüglich der Problematik der Erzeugung von weißem Rauschen sei auf Kapitel 6 verwiesen.) Abbildung 2.2 (a) zeigt das Ergebnis für eine Fernleitung, wie deutlich an dem Abbruch des Spektrums bei 3400 Hz zu erkennen ist. Die Abbildungen 2.2 (b) und 2.2 (c) zeigen die Werte für eine Ortsleitung bzw. eine hausinterne Verbindung. Es ist deutlich zu erkennen, z. B. an dem Minimum bei 2500 Hz, daß die drei Spektren - abgesehen von dem Fehlen der Frequenzen oberhalb von 3400 Hz im Ferngespräch - im wesentlichen denselben Verlauf aufweisen. Daraus ist zu entnehmen, daß nicht die Fernsprechleitung maßgeblich das Spektrum prägt, sondern daß dies durch die Telefonapparate, und hier mit großer Wahrscheinlich-

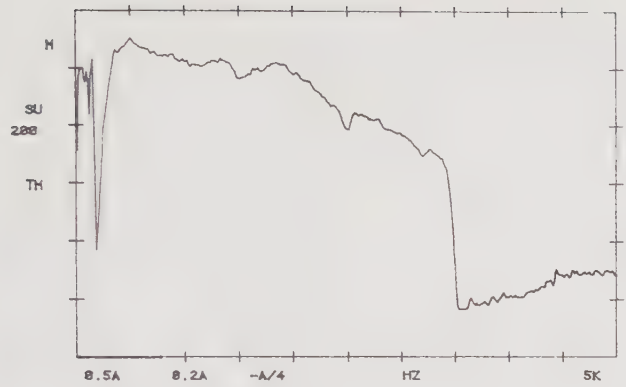


Abb. 2.2 (a):
Fernleitung
Sprechkapsel I

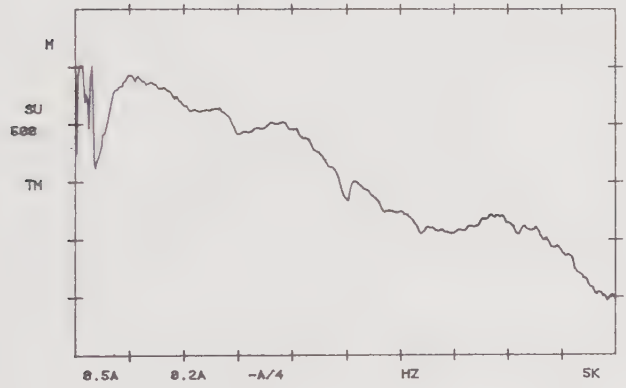


Abb. 2.2 (b):
Ortsleitung
Sprechkapsel I

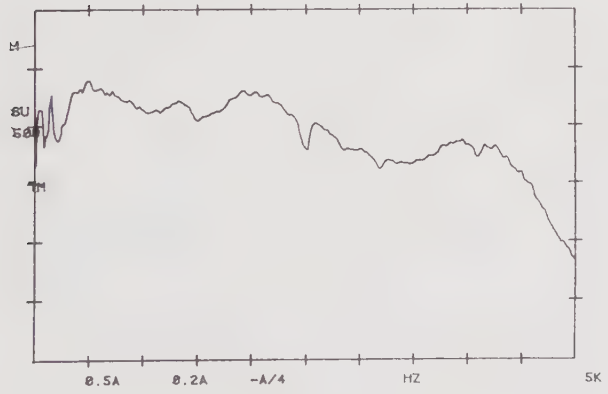


Abb. 2.2 (c):
Hausleitung
Sprechkapsel I

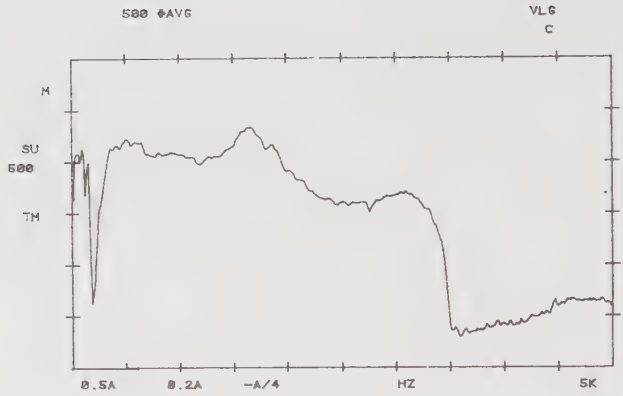


Abb. 2.3 (a):
Fernleitung
Sprechkapsel II

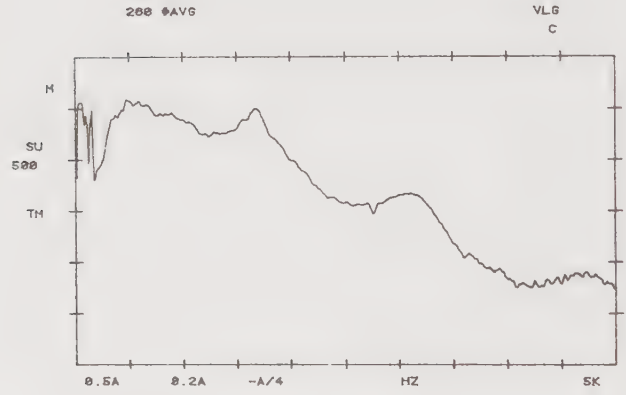


Abb. 2.3 (b):
Ortsleitung
Sprechkapsel II

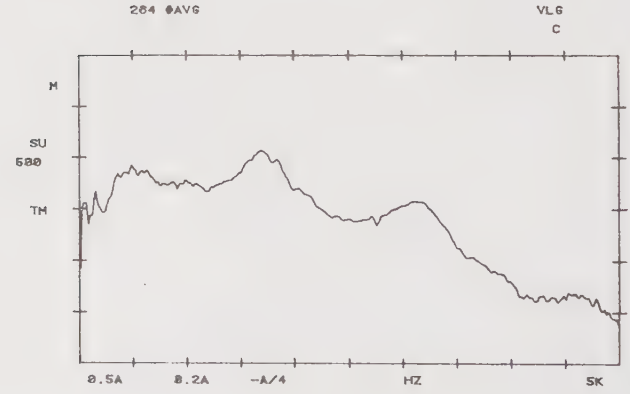


Abb. 2.3 (c):
Hausleitung
Sprechkapsel II

keit durch die Sprechkapsel geschieht. Unterstützt wird dieses Ergebnis durch die Spektren in den Abbildungen 2.3 (a-c), die sich von den Messungen in den Abbildungen 2.2 (a-c) nur dadurch unterscheiden, daß die Sprechkapsel ausgetauscht wurde. Abbildung 2.3 (a) bezieht sich also auf den identischen Leitungsweg wie Abbildung 2.2 (a) usw. Wieder ist zu beobachten, daß die Spektren für Fern-, Orts- und Hausleitung einen sehr ähnlichen Verlauf aufweisen. Auffällig ist auch die besonders stark strukturierte Ausprägung der Spektren in den Abbildungen 2.3 (a-c). Die Ursache ist offensichtlich, daß hier eine alte und seit Jahren in Gebrauch befindliche Kapsel benutzt wurde, während die Messungen in den Abbildungen 2.2 (a-c) mit einer fabrikneuen Kapsel durchgeführt wurden.

2.2 Möglichkeiten der Kanalkompensation

Jeder der vorgenannten Störungseinflüsse kann mit modernen Methoden der Signalverarbeitung bis zu einem gewissen Grad reduziert werden. Diese Aktivitäten, die unter dem Begriff "Sprachverbesserung" zusammengefaßt werden, stellen allein schon wegen ihrer großen praktischen Bedeutung heute ein eigenes Gebiet dar. Entsprechend viele Veröffentlichungen mit zum Teil ausführlichen Übersichtsdarstellungen sind zugänglich, von denen hier nur einige mit besonderem forensischen Bezug erwähnt werden sollen (LIM 1978, BEEK 1977, LEWIS 1980, BISIGNANI, u.a.) Grundsätzlich muß unterschieden werden, welchem Ziel die Sprachverbesserung dienen soll. Es kann die Verbesserung der Verständlichkeit (Intelligibility), der Zuhörbarkeit (Listeneability) oder auch eine Verbesserung als Vorverarbeitungsstufe zu einer folgenden technischen Signalverarbeitung, z. B. der Ableitung stabiler sprecher- oder lautspezifischer Merkmale durch Computer, angestrebt werden. Im letzteren Fall ist die "Erfolgskontrolle" erheblich einfacher als in den zuerst genannten, da algorithmische Unterschiede meßmethodisch relativ leicht erfaßbar sind, während bei der Sprachverständlichkeit

die Komplexität der menschlichen Perzeption eingeht und Messungen (z.B. durch sogenannte "Diagnostic Rhyme Tests") daher in ihrer Aussagekraft für Anwendungen schwer zu bewerten sind.

Ein besonders schwieriges Problem in diesem Zusammenhang, das hier nur erwähnt werden kann, ist die Trennung von Sprache verschiedener gleichzeitig sprechender Personen. Einfache Charakteristika z. B. zur Trennung von Sprache und Hintergrundgeräuschen können hier nicht benutzt werden, da Sprache von Sprache zu unterscheiden ist. Einzelheiten hierzu finden sich in PARSONS (1978).

Bisher sind in der Literatur keine Ansätze bekannt geworden, die die Ableitung von Übertragungseigenschaften des Telefonkanals aus dem bereits übertragenen Sprachsignal zum Ziel hatten. Dagegen gibt es viele Versuche, ihren Einfluß auf relevante Parameter zu reduzieren ohne den Kanal selbst zu betrachten. Dabei wird oft davon ausgegangen, daß Mittelwerte von Meßgrößen über längere Zeitsegmente nicht nur mittlere Eigenschaften des Sprechers bzw. der Sprachlaute, sondern auch der Übertragungseigenschaften des Kanals enthalten müssen. Werden nun Kurzzeitmeßgrößen - die ja lautbezogene Abweichungen von den Mittelwerten aufweisen - durch Langzeitwerte dividiert, müßten sich die Kanaleigenschaften automatisch kompensieren lassen, ohne daß sie im einzelnen bekannt werden. So vielversprechend dieser Ansatz zunächst scheint, bisher sind keine praktischen Anwendungen bekannt geworden, obwohl daran gearbeitet wurde. Andere Verfahren ermitteln die Eigenschaften der Hintergrundgeräusche in den Sprechpausen und restaurieren damit das Sprachsignal.

Das in dieser Arbeit beschriebene Verfahren hat die Ermittlung der Kanaleigenschaften selbst zum Ziel. Insofern stellt es zu den genannten Verfahren eine Ergänzung dar. Möglichkeiten seines Einsatzes zur Kanalkompensation ergeben sich dann zwangsläufig aus der Kenntnis der Übertragungseigenschaften.

2.3 Kanalvermessung mithilfe von Sprachlautspektren

Die Frequenzeigenschaften H(w) einer Übertragungsstrecke lassen sich unter bestimmten Voraussetzungen gemäß

$$F(w) = H(w) * Q(w) \quad \text{mit} \quad w = 2\pi/T$$

aus dem empfangenen Signalspektrum F(w) wiedergewinnen, wenn das Quellenspektrum S(w) bekannt ist. Dies ist beim Einsatz von Sprachverabeitungssystemen für Telefonsprache im allgemeinen nicht der Fall, da aus Gründen der Praktikabilität (kommerzieller Bereich) oder wegen der nicht mehr bestehenden Verbindung (forensischer Bereich) kein Signal zur Messung (im folgenden auch Eichsignal genannt) übertragen werden kann. Es liegt daher nahe, im Sprachsignal selbst nach Lautelementen zu suchen, die sich durch möglichst große Invariablität gegenüber verschiedenen Sprechern, Kontexten und Realisationen, auszeichnen und somit nach vorheriger Ausmessung als Eichsignal dienen könnten. Dazu wurde von GOYDKE (1973) vorgeschlagen, Frikative heranzuziehen, die - wie bereits geschildert - durch ihren weit vorn liegenden Artikulationsort nur wenig sprecherspezifische Information aufnehmen können. Speziell genannt werden von GOYDKE /f/ und /s/, wobei er bereits darauf hinweist, daß das Spektrum des /s/ weitgehend sprecherunabhängig, aber nicht unbedingt im Frequenzverlauf konstant ist. Der Vorzug des /s/ bestünde in seiner größeren Lautstärke gegenüber dem /f/, das um ca. 15 dB unter dem mittleren Pegel der gesamten Lautstärke liegt und daher bei Sprachaufzeichnungen mit geringem Störspannungsabstand möglicherweise als Eichsignal nicht geeignet ist. Die eigene Erfahrung in der forensischen Sprachanlyse zeigt außerdem, daß die /s/-Spektren deutliche sprecherspezifische Ausprägungen aufweisen und von daher als Eichsignal nicht verwendet werden können. So ist beispielsweise bei Prothesenträgern im höherfrequenten Bereich oft ein deutliches Energiemaximum zu beobachten, das auch als Pfeifton hörbar ist.

Es erscheint daher sinnvoll, durch eine Analyse der akustischen Struktur des Frikativs /f/ ein Verfahren zu entwickeln, daß seine Benutzung als Eichgröße erlaubt. Dabei beziehen sich die folgenden Überlegungen ausschließlich auf die Ermittlung der Frequenzeigenschaften der Übertragungsstrecke. Im Hinblick auf eine forensische Auswertung (sowohl auditiv-phonetisch wie auch als Vorstufe für maschinelle Messungen) ist dies die wichtigste Eigenschaft, denn Frequenzinformationen werden sowohl vom Menschen wie auch von allen Sprach- und Sprechererkennungssystemen verwendet. Voraussetzung für die Gültigkeit der folgenden Betrachtungen ist allerdings, daß keine wesentlichen nichtlinearen Verzerrungen vorliegen, wie sie z.B. bei Verwendung von Kohlekapseln als Mikrophon auftreten können. Andererseits könnten im Falle einer Anwendung im kommerziellen Bereich (z.B. nach einer Automatisierung der Segmentation als Vorverarbeitung in der Spracherkennung) leicht dynamische Mikrophone verwendet werden. Bezüglich der forensischen Anwendungsmöglichkeiten ist anzumerken, daß sämtliche Münzfernsprecher in der Bundesrepublik Deutschland, von denen die große Mehrzahl der Anrufe eingeht, mit dynamischen Mikrofonen ausgestattet sind. Außerdem führen mäßige nichtlineare Verzerrungen nicht zu einer völligen Unbrauchbarkeit der Messung, sondern reduzieren deren Genauigkeit.

2.4 Zum Gegenstand der vorliegenden Untersuchung

Die Berechnung der Spektren kann auf modernen Rechenanlagen mittels Fouriertransformation (realisiert als Fast-Fourier-Transformation, FFT, siehe Kapitel 4) durchgeführt werden. Die Parameter zur Festlegung der FFT müssen so gewählt werden, daß die daraus resultierenden Größen wie Auflösung, Bandbreite etc. der akustischen Struktur der Frikative angemessen sind. Sodann ist zu untersuchen, ob und wie sich /f/-Spektren bezüglich verschiedener Sprecher, Kontexte und Realisationen verändern. Dazu werden sie in verschiedener Weise durch Mittelung zusammengefaßt:

Typ 1: Mittelung über 1 Sprecher, 1 Realisation, alle Kontexte

Typ 2: Mittelung über 1 Sprecher, alle Realisationen, 1 Kontext

Typ 3: Mittelung über 1 Sprecher, alle Realisationen, alle Kontexte

Typ 4: Mittelung über alle Sprecher, alle Realisationen, 1 Kontext

Typ 5: Mittelung über alle Sprecher, Realisationen und Kontexte

Für die spätere Anwendung wird diejenige Klasse von Spektren am besten geeignet sein, die die geringsten Streuungen aufweist. Um diese möglichst gering zu halten, können neben einer geeigneten Festlegung der Parameter bei der Berechnung der Spektren verschiedene Wege bei der Auswahl des der Berechnung zugrunde liegenden Signalsegmentes gegangen werden. Nach Segmentation des /f/ bieten sich nämlich die folgenden Möglichkeiten:

- Für jede /f/-Realisierung wird genau ein Spektrum berechnet, wobei das Zeitintervall für die Berechnung in die Mitte des /f/-Segmentes gelegt wird, um den kontextuellen Einfluß möglichst gering zu halten.

- Für jede /f/-Realisierung wird das Fenster für die FFT mit einem vorgegebenen Überlappungsgrad vom Beginn des Lautes bis zum Ende verschoben und die sich ergebenden Spektren gemittelt. Für die weiteren Untersuchungen werden dann diese Mittelwertspektren herangezogen.

Mit der an zweiter Stelle genannten Methode stellt sich auch - zwangsläufig die allgemeinere phonetische Frage nach den intrinsischen und den kontextuell bedingten Dauern der einzelnen /f/-Realisierungen, aus denen sich die Zahl der möglichen Mittelungen ergibt. Hierauf wird in Abschnitt 5.1 näher eingegangen.

Vor der Weiterverarbeitung sind die Spektren zu normieren, z.B. auf gleiche Fläche oder, wie in dieser Arbeit geschehen, auf denselben Mittelwert. Eine weitere Minimierung der Variabilität der Spektren kann möglicherweise dadurch erreicht werden, daß nicht über den gesamten Frequenzbereich normiert wird, sondern daß Normierung und evtl. auch Auswertung auf den Telefonkanal beschränkt werden.

Das schwierigste Problem der gesamten Untersuchung ist zweifellos die Erfassung der Streubreiten, d.h. die quantitative Ermittlung der Amplitudenverteilungen innerhalb der einzelnen Frequenzkanäle in Abhängigkeit von der Anzahl der gemittelten Spektren. Eine solche Beschreibung ist für die forensische Anwendung unverzichtbar, da in jedem Einzelfall ein Maß für die Sicherheit angebbar sein muß, mit der eine Messung (bzw. spätere Entscheidung) vorgenommen wird. Auf dieses Problem wird in Abschnitt 5.3 ausführlich eingegangen. Schließlich ist in praxisgerechten Versuchen der Nachweis zu erbringen daß das beschriebene Verfahren für den Einsatz im forensischen Bereich geeignet ist. In Kapitel 6 werden verschiedene Anwendungsbeispiele aufgeführt.

3. DIE SPEKTRALE STRUKTUR DER /f/-LAUTE

Im Vergleich zu Okklusiven sind Frikative relativ wenig untersucht worden. Dies gilt in besonderem Maße für das /f/, das wegen seiner fehlenden Ergiebigkeit als Merkmalsträger und seiner zumindest zu vermutenden strukturellen Unergiebigkeit keinen "attraktiven" Forschungsgegenstand darstellt. Dennoch finden sich in der Literatur Messungen wie auch theoretische Modellbetrachtungen zur Spektralstruktur des /f/, die neben eigenen Überlegungen in diesem Kapitel dargestellt werden.

Diese Arbeit enthält zahlreiche Meßergebnisse, die sich auf die notwendigen Voruntersuchungen, die Struktur des Frikativs /f/ und die Nutzung im forensischen Bereich beziehen. Der Übersichtlichkeit wegen und um direkte Vergleiche mit den Ergebnissen anderer Autoren vornehmen zu können, hat es sich bei der Fülle des Materials als sinnvoll erwiesen, die Meßergebnisse nicht an späterer Stelle insgesamt darzustellen, sondern sie bereits in diesem Kapitel - soweit notwendig - den jeweiligen theoretischen Überlegungen hinzuzufügen bzw. den Ergebnissen aus der Literatur gegenüberzustellen.

3.1 Akustische Überlegungen

In der durch FANT (1970) begründeten akustischen Theorie der Sprachproduktion werden die Schallquelle, das durch das Ansatzrohr gebildete akustische Filter sowie die Abstrahlverluste am Mund als wesentliche Komponenten angesehen. Für Frikative bilden in einem sehr einfachen Modell die durch die Enge im Vokaltrakt hervorgerufenen Turbulenzen die Schallquelle und der verbleibende Rest des Vokaltraktes bis zu den Lippen das akustische Filter. Bei allen Modellrechnungen auf dieser Basis stellt die akustische Terminierung des Ansatzrohres an den Lippen ein besonderes Problem dar, das sich aufgrund der Größenverhältnisse umso

fehlerhafter auswirkt je kürzer das Ansatzrohr ist. Hier wird die Problematik von Modellrechnungen für das /f/ bereits deutlich, das ja überhaupt erst an den Lippen gebildet wird. Nach UNGEHEUER (1962) sind die Resonanzfrequenzen F_n eines kreiszylindischen, einseitig geschlossenenen Rohres unter Vernachläßigung aller sonstigen Effekte, insbesondere auch der Abstrahlverluste, gegeben durch

$$F_n = \frac{(2\,n - 1)\,c}{4L}$$

wobei c die Schallgeschwindigkeit und L die Länge des Ansatzrohres ist. Mit diesen extrem vereinfachten Annahmen ergeben sich bei einer Schallgeschwindigkeit von c=340 m/sec und einer angenommenen Länge des Ansatzrohres für ein /f/ (praktisch die Länge der Lippenstülpung) von

L = 1 cm:

F_1 = 8500 Hz und F_2 = 25500 Hz

und

L = 2 cm:

F_1 = 4250 Hz und F_2 = 12750 Hz

Ähnliche Rechnungen finden sich in der Literatur bei HEINZ und STEVENS (1961), die in ihrer Arbeit für den ersten Formanten von der Halbwellenresonanz eines 2 cm - Rohres ausgehen und dafür 8000 Hz angeben. Dieser Betrachtung liegt allerdings ein beidseitig offenes Rohr, entsprechend anderen Randbedingungen für die Differentialgleichung, von der diese Rechnungen ausgehen, zugrunde. Insgesamt zeigt dieses Beispiel, daß Modellrechnungen bei den für die Bildung von /f/-Lauten gegebenen artikulatorischen Konfigurationen aufgrund der sensiblen Abhängigkeit entscheidender, aber nicht hinreichend exakt bekannter oder zu schätzender Parameter äußerst problematisch sind.

Die ausführlichste Theorie zur Produktion der Frikative findet sich nach wie vor bei FANT (1970). Wie bereits erwähnt wirkt bei

Frikativen die an der verengten Stelle im Ansatzrohr auftretende turbulente Strömung als Schallquelle. Aus der Strömungsphysik ist bekannt, daß bestimmte Bedingungen erfüllt sein müssen, damit eine laminare (gleichförmige) Luftströmung in eine turbulente übergeht, nämlich dann, wenn die dimensionslose Reynold'sche Zahl, in die als maßgebliche Größen die Beschaffenheit des Strömungsmediums (hier der Luft), der wirksame Querschnitt der Enge sowie die Strömungsgeschwindigkeit eingehen, einen Schwellwert überschreitet. Eine turbulente Strömung kann von Sprechern an praktisch jeder Stelle des Vokaltraktes erzeugt werden. Solche unterschiedlichen Artikulationsstellen können allophonisch (z.B. [x] und [ç]) oder phonematisch relevant sein. Für das hier betrachtete /f/ des deutschen Lautsystems entsteht die Enge durch Anlegen der Unterlippe an die oberen Schneidezähne (labiodental). Einige Angaben zu den hierfür geltenden physikalischen Gegebenheiten (Druckverhältnisse etc.) finden sich bei MEYER-EPPLER (1953) und FANT (1970).

Die Annahme einer regellosen Signalquelle als Folge einer turbulenten Strömung ist sicher nicht unproblematisch, wie schon TARNOCZY (1954) erkannt hat. Er schreibt über die Streuungen von Frikativspektren:

> "Die meisten Störungen bedeuten die momentanen Schwankungen der Phonation, wie z.B. die bei der Aussprache von /f/ und /θ/ von den die Spalte teilweise immer wieder verschließenden, und bei ihrem jeweiligen Aufspringen wieder freigebenden, Speichelblasen verursachten Schwankungen."

Ähnliche Effekte werden von FANT für Okklusive beschrieben. Er schließt zudem die Möglichkeit nicht aus, daß der Luftstrom auch durch vereinzelte Vibrationen der Stimmlippen moduliert wird (siehe hierzu Abschnitt 3.2)

Von besonderem Wert für das Verständnis der physikalischen Vorgänge bei der Produktion der Frikative sind die vergleichenden

Untersuchungen FANTs mit Daten aus einem elektrisch nachgebildeten akustisch-phonetischen Modell und echten Meßergebnissen. FANTs Berechnungen stützen sich zunächst auf möglichst genaue Nachbildungen des Vokaltraktes anhand von Querschnittsfunktionen, die aus Röntgendaten gewonnen wurden. Sein Modell berücksichtigt ebenfalls Abstrahlverluste am Mund sowie bei stimmlosen Frikativen die dämpfende Wirkung der weit geöffneten Glottis auf die Formanten. Letzteres führt neben einer Dämpfung zu einer Erhöhung des ersten Formanten. Er erreicht mit diesem Modell eine außerordentlich hohe Übereinstimmung von Meß- und Modelldaten (Einzelheiten siehe Abschnitt 3.2).

In einem weiteren Modell versucht FANT, die komplizierte Form des Ansatzrohres zunächst durch ein vereinfachtes gestuftes Rohr, dann durch ein Doppelrohr als Zweifachresonator, zu ersetzen. Dieses Modell, daß in besonderem Maße für Frikative der physikalischen Wirklichkeit entgegenkommt, erlaubt auf vergleichsweise einfache Art, die Einflüsse beider Resonanzräume, nämlich vor und hinter der Enge, im Detail zu studieren.

FANT unterscheidet bei den labial und labio-dental gebildeten Lauten zwischen nichtpalatalisierten mit dem Hauptresonator am vorderen Gaumen und den palatalisierten mit dem Resonanzraum im hinteren Rachenraum. Bei den /f/-Lauten unterscheidet er ein nichtpalatalisiertes Allophon [f] und ein palatalisiertes [f,]. Er kann an einem einfachen Resonatormodell nachweisen, daß die Palatalisierung zu einer Verstärkung des zweiten und dritten Formanten sowie zu einer Antiresonanz vor F_2 führt.

3.2 Messungen zur Spektralstruktur

3.2.1 /f/-Spektren in lautlichem Kontext

Zentraler Gegenstand in der Beschreibung von Spektralstrukturen ist in der Regel die Lage und Form der Formanten. TARNOCZY (1954) wirft die Frage auf, ob dieser Begriff, ursprünglich als Energiekonzentration im Spektrum der Vokale gebräuchlich, für Frikative ohne vergleichbar ausgeprägte Formantstruktur überhaupt anwendbar sein kann:

> "Der Formant ist hinsichtlich seines Frequenzgebietes dermaßen beschränkt und solcherart aus seiner Umgebung hervorspringend, daß er durch die Frequenzstelle seines Kerns eindeutig gekennzeichnet werden kann. Durch die zwei Hauptformanten dargestellt, werden die Vokale von einander klar abgegrenzt, wobei die Formantangaben sozusagen auch die Bildungsstelle und die Bildungsart der Vokale widerspiegeln. Zur selben Zeit erscheinen die hervorspringenden Frequenzstellen der Geräuschlaute breit und unbestimmt. Von der Frequenzstelle irgendeines Kerns kann gar nicht gesprochen werden, ist doch das Frequenzspektrum im ganzen Hörbereich fast gleichmäßig. Hinzukommen noch die Streuungen der individuellen Aussprache, um die hervorspringenden Frequenzstellen noch weiter zu egalisieren."

TARNOCZYs Bedenken werden in dieser Arbeit nicht geteilt, obgleich sicherlich anzuerkennen ist, daß niemand ein flaches, über das gesamte hörbare Spektrum verteiltes Energiemaximum als Formant bezeichnen wird. Im übrigen zeigen Messungen (siehe weiter unten), daß das Spektrum der "Geräuschlaute" durchaus Verläufe aufweisen kann, die zweifelsfrei die Angabe definierter Frequenzwerte als Formantpositionen erlauben.

Bezüglich der Eigenschaften der /f/-Spektren ist wieder zunächst ein Blick auf die Arbeiten von FANT geboten. Mit einer Filterbank bzw. einem Filter mit variabler Mittenfrequenz wurden die Spektren der einsilbigen Testwörter /fa/ und /f,a/ bestimmt. Wegen ihrer Bedeutung für die späteren Vergleiche in dieser Arbeit werden in den Abbildungen 3.1 und 3.2 die Spektren für das nichtpalatalisierte /f/ und das palatalisierte /f,/ gezeigt, entnommen (FANT 1970).

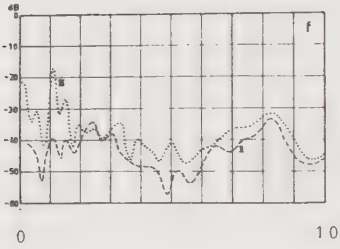


Abb.3.1

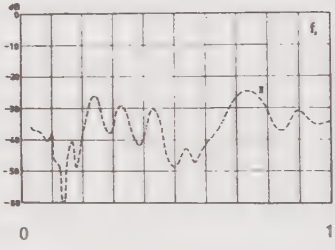


Abb. 3.2

Spektren des nichtpalatalisierten /f/ bzw. des palatalisierten /f,/ nach FANT

Abb. 3.1 für /f/ zeigt zwei Spektren, die zeitlich an verschiedenen Stellen des Signals gewonnen wurden. Die mit III bezeichnete Kurve liegt im Vergleich zu den in beiden Abbildungen mit II gekennzeichneten Kurven im Zeitsignal näher am Einsatzpunkt der Phonation zum /a/. FANT bezeichnet den Verlauf der Spektren als im wesentlichen flach. Dies trifft sicherlich zu, wenn der Gesamtverlauf z.B. mit Vokalspektren verglichen wird. Andererseits ist doch eine ausgeprägte Struktur mit um mehr als 20 dB auseinanderliegenden Extremwerten zu beobachten, so daß für die Zwecke dieser Arbeit wohl kaum von einem ausgeglichenen Spektrum

gesprochen werden kann.

In beiden Kurven II fallen besonders die Minima bei 750 Hz in [f] bzw. 13000 Hz in [f,] auf. Hauptresonanzen sind die Maxima bei 8500 Hz bzw. 7500 Hz, hervorgerufen durch den Resonanzraum an den Lippen. Die dazwischenliegende Struktur erklärt FANT mit dem Vorliegen weiterer Schallquellen, die neben den bereits genannten vereinzelten Glottisschlägen insbesondere bei palatalisiertem [f,] in zusätzlichen Geräuschquellen an der durch die Palatalisierung entstehenden Enge liegen. Bemerkenswert ist auch die Beobachtung FANTs, daß (nur aus einem zusätzlichen, hier nicht wiedergegebenen Spektrogramm zu ersehen) im Beispiel III F_2 (das Maximum bei ca. 1000 Hz) unmittelbar vor der Phonation zum /a/ höher liegt als der Beginn der F_2-Transition innerhalb des folgenden Vokals. FANT nimmt dies als experimentelle Bestätigung für die bereits genannte Formantverschiebung zu höheren Frequenzen bei offenstehender Glottis, jedenfalls solcher Formanten, die durch den Pharynxraum entstehen.

Neben FANT haben sich auch andere Autoren mit der Struktur der Frikative befaßt. Die älteste Arbeit mit quantitativen Angaben scheint die Untersuchung von HERMANN (1894) zu sein, der für /f/ Formantwerte von 1400 bis 1550 Hz und 1750 bis 2050 Hz angibt. Ähnliche Angaben macht STUMPF (1926), der mit seinem um die Jahrhundertwende entstandenen Interferenzröhrenapparat "charakteristische Frequenzstellen" bei 1400 Hz - 2500 Hz ermittelte. Verglichen mit den Spektren FANTs sind diese Werte, insbesondere für das palatisierte [f,] für die damaligen Meßmethoden erstaunlich wirklichkeitsgetreu. Nach Angaben von TARNOCZY (1954) wurden 1934 von GEMELLI und PASTORI durch Auszählen der Perioden Formanten des /f/ (aus den italienischen Wörtern /finiamo/ und /faggio/) zu 5500 Hz und 4200 Hz bestimmt. Diese Werte fallen nun allerdings bezogen auf Angaben anderer Autoren in einen Spektralbereich minimaler Energie und sind wohl als generelle Charakterisierung schlicht unzutreffend. TARNOCZY selbst findet (1954) ebenfalls durch Auszählen Formanten bei 7200 Hz, ein Wert, der sehr gut mit FANTs Daten sowie theoretischen Werten überein-

stimmt. Allerdings anerkennt TARNOCZY in der zitierten Arbeit nicht die Existenz von /f/-Allophonen oder auch nur die Beeinflussung des /f/ durch koartikulatorische Effekte.

Die Auffassung, daß die Spektren von /f/-Lauten als flach anzusehen sind, wird von TARNOCZY (1971) geteilt.

> "Der Spektrum-Pegel ist mit einer Genauigkeit von +/- 1 dB zwischen 200 und 16000 Hz konstant. Dies bedeutet, daß man so einen Generator für weißes Rauschen mit endlosem Band am einfachsten und billigsten erzeugen kann."

Abgesehen davon, daß dies in technischer Hinsicht schon lange nicht mehr stimmt, ist die Angabe +/- 1.5 dB erstaunlich, steht sie doch im Widerspruch zu allen anderen Autoren, sogar zu seiner eigenen Arbeit aus dem Jahre 1954. Sein Meßverfahren bestand in der sukzessiven Anwendung von Filtern auf Sprachaufzeichnungen, die er mit seiner "Sprechchormethode" (TARNOCZY 1970) gewonnen hatte, bei der mehrere Sprecher gleichzeitig sprechen. Die Annahme ist dabei, daß dadurch von vornherein eine Art von Mittelung entsteht. TARNOCZY behauptet sogar, koartikulatorische Effekte für /f/, die er in der früheren Arbeit verneinte, dadurch eliminieren zu können, daß er mehrere Personen gleichzeitig Wörter aussprechen läßt, die den aktuellen Frikativ an zeitlich derselben Stelle enthalten (Beispiel /lassen/, /Hessen/, /Klassik/), und so zu einem "abstrakten" Laut zu kommen. Die Probleme und Ungenauigkeiten, die dabei entstehen, sind offensichtlich.

K. BRAHM (1950) veröffentlichte Ergebnisse von Messungen mit dem FREYSTEDTschen Tonfrequenzspektrometer an elf Versuchspersonen. Für das /f/ ermittelte er ein Hauptmaximum bei 5000 - 12000 Hz; ein nicht sehr vielsagendes Ergebnis.

Zur spektralen Struktur der /f/-Laute schreibt TSCHESCHNER (1964), daß das /f/ "lediglich durch ein gleichmäßiges verdeckendes Rauschen dargestellt wird". Dabei geht er auch auf die in perzeptiver Hinsicht bedeutungsvolle Bandbreite ein. So wird /f/ bei einem Hochpaß mit der Grenzfrequenz f_g = 2.5 kHz (Im Original findet sich die Angabe 2.5 Hz, offensichtlich ist aber 2.5 kHz gemeint.) als /ʃ/ und bei einem Tiefpaß mit f_g = 1.1 kHz als /h/ empfunden (TSCHESCHNER 1964). Nach TSCHESCHNER ist für den subjektiven Erkennungsprozeß von Frikativen nicht nur die relative sondern auch die absolute Lage ihrer Formanten von Bedeutung. Diese Erkenntnis ist im Grunde trivial, da bei entsprechender Verschiebung im Frequenzbereich sich zwangsläufig irgendwann ein Spektrum ergibt, das demjenigen eines anderen Sprachlautes ähnlich ist. Für bestimmte Frikative folgt daraus allerdings, daß bei flacher werdendem Spektrum aufgrund äußerer Einflüsse auf das Sprachsignal, z.B. durch die Kohlekapsel des Telefonapparates, der Höreindruck zum /f/ tendiert. Nach TSCHESCHNER wird beispielsweise /s/ bei einem Verhältnis von verschobener zur Originalfrequenz von 1:3 als /f/ empfunden. TSCHESCHNER erklärt diesen Effekt damit, daß der Formant des /s/, den er zwischen sechs und neun kHz angibt, dann nicht mehr in den hörbaren Bereich fällt. Bei telefonübertragener Sprache, die ja wesentlich für die Untersuchungen dieser Arbeit ist, können ähnliche Effekte allein schon durch die Bandbegrenzung bei 3.4 kHz in der Fernebene oder auch durch zufällige kompensatorische Übertragungseigenschaften des Gesamtkanals verursacht werden.

HEINZ und STEVENS (1961) geben in ihrer Arbeit über stimmlose Konsonanten für /f/-Realisationen aus den Wörtern /aloof/, /cuff/ und /fear/ Spektren an, die in der Abbildung 3.3 dargestellt sind. Die gepunkteten Linien sind gemessene Spektren, die durchgezogenen Kurven wurden aus den bereits genannten Modellrechnungen ähnlich denen FANTs gewonnen. Ein Vergleich mit den Spektren in Abbildung 3.1 und 3.2 zeigt, daß das Hauptmaximum zwischen 7000 und 10000 Hz liegt, wie es die akustische Theorie auch erwarten läßt.

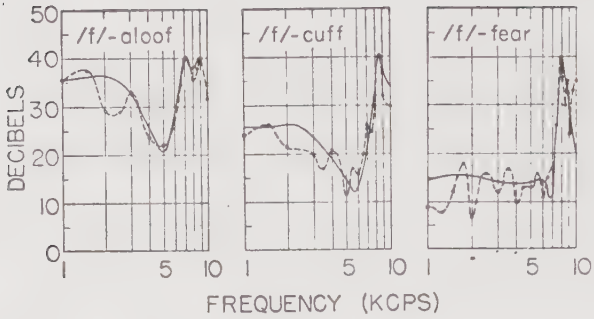


Abb. 3.3: /f/-Spektren aus verschiedenen Kontexten, entnommen STEVENS & HEINZ (1961).

Die Spektren aus /aloof/ und /cuff/ entsprechen wohl am ehesten dem nichtpalatalisierten /f/ bei FANT und zeigen auch - soweit es die unterschiedlichen Frequenzskalen zulassen - ähnliche Struktur, wobei besonders auffallend die exakte Übereinstimmung des Minimums bei 5000 Hz ist. Auch für /f/ aus /fear/ zeigt sich, wenn auch nicht im Detail, gute Übereinstimmung. Bedauernswert für die Zwecke dieser Arbeit sind die fehlenden Angaben bei HEINZ und STEVENS für den Bereich unter 1000 Hz.

3.2.2 /f/-Spektren in Isolation

Wie FANTs Messungen bereits andeuten, sind die spektralen Eigenschaften des /f/ keineswegs kontextunabhängig. Da aber andererseits aus phonetischer Sicht die Kenntnis der Struktur eines "kontextlosen" /f/, artikulatorisch gesehen des /f/-Targets, von besonderem Interesse ist, liegt nahe, Spektraluntersuchungen an

isoliert produzierten /f/-Realisierungen vorzunehmen. Dabei zeigte sich, daß sämtliche Versuchspersonen zunächst das isolierte /f/ mit viel zu hoher Intensität realisierten. Die von TARNOCZY beschriebenen "Speichelblasen" (siehe Abschnitt 3.1) waren dadurch fast regelmäßig zu beobachten. Erst nach einiger Gewöhnung der Sprecher konnten die isolierten /f/-Realisationen anhaltend mehrere Sekunden produziert und mit etwa der Intensität aufgenommen werden, wie sie auch in natürlicher Sprache vorliegt. Die Analyse erfolgte mit einem FFT-Analysator, mit dem in Realzeit das Leistungsdichtespektrum berechnet werden konnte.

Zunächst wurde geprüft, ob sich eine Veränderung des Spektrums ergibt, wenn ein Mikrophon mit bzw. ohne Schaumstoff-Windschutz benutzt wird. Die Abbildungen 3.4 und 3.5 zeigen das Ergebnis bis 10 kHz. Die Form des Spektrums bleibt erhalten, lediglich geringfügige Abweichungen in der Amplitude sind an manchen Stellen zu beobachten, die jedoch mit einiger Sicherheit auf die unterschiedlichen Realisationen zurückzuführen sein dürften.

Abbildung 3.6 zeigt das Spektrum eines isoliert gesprochenen /f/ eines zweiten Sprechers. Es ergeben sich geringfügige Abweichungen in der Amplitude, jedoch ist im wesentlichen dieselbe Formantstruktur zu beobachten.

Die Abbildungen 3.7 bis 3.8 zeigen weitere Spektren von isoliert gesprochenen /f/, die alle von derselben Versuchsperson stammen. Der Verlauf der Spektren in der Feinstruktur ist glatter, da die Zahl der Mittelungen über aufeinander folgende einzelne Spektren von 40 auf über 100 erhöht wurde. Dadurch läßt sich die Lage der Minima und Maxima sicherer ermitteln. Untersucht wurden in Anlehnung an die Arbeiten von FANT die /f/-Allophone [f] und ein palatalisiertes [f,]. Abbildung 3.7 zeigt das Spektrum für [f] bis 20 kHz. Eindeutig ist ein ausgeprägter Formant bei 9200 Hz zu erkennen (markiert durch die Cursorposition). Dieser Formant ist zweifelsfrei derjenige, der sich auch bei akustischen Modellrechnungen ergibt (erste Halbwelle in der akustisch-phonetischen Theorie, siehe Abschnitt 3.1). Seine Lage stimmt

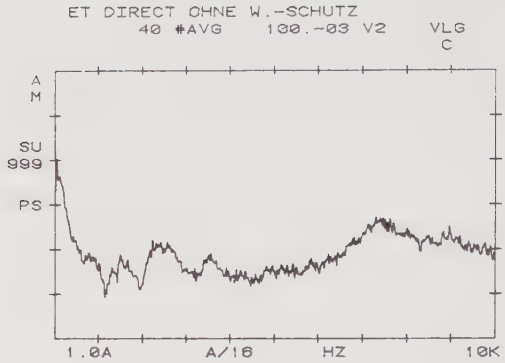


Abb. 3.4: Spektrum eines isoliert gesprochenen /f/, Sprecher 1, ohne Windschutz

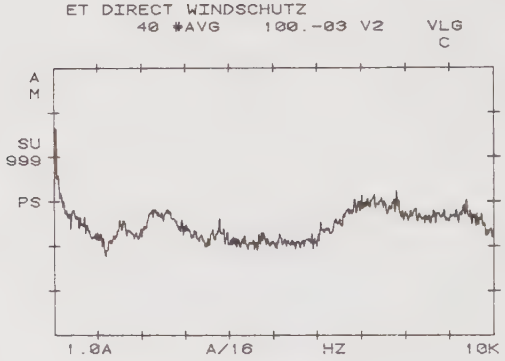


Abb. 3.5: Spektrum eines isoliert gesprochenen /f/, Sprecher 1, mit Windschutz

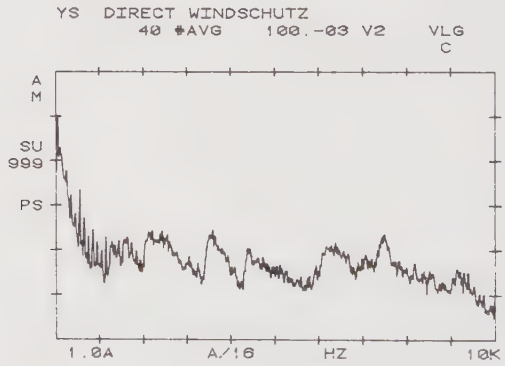


Abb. 3.6: Spektrum eines isoliert gesprochenen /f/, Sprecher 2, mit Windschutz

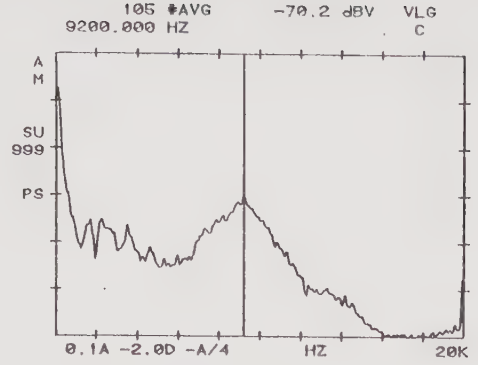


Abb. 3.7: Isoliert gesprochenes [f], Sprecher 1, Realisation 1

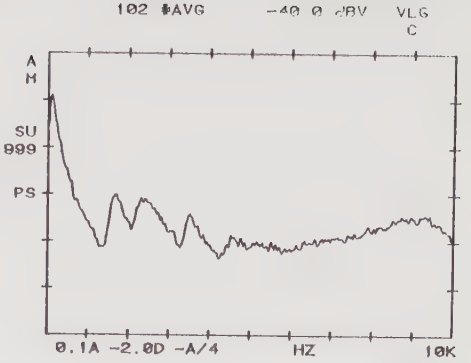


Abb. 3.8: Isoliert gesprochenes [f], Sprecher 1, Realisation 2

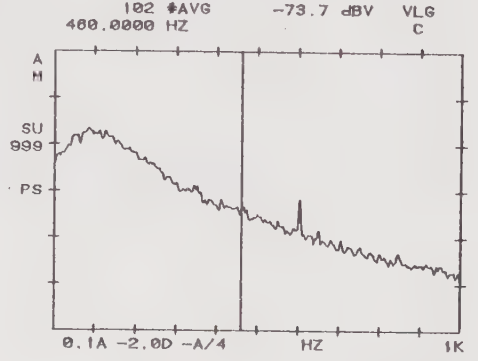


Abb. 3.9: Isoliert gesprochenes [f] Sprecher 1, Realisation 3

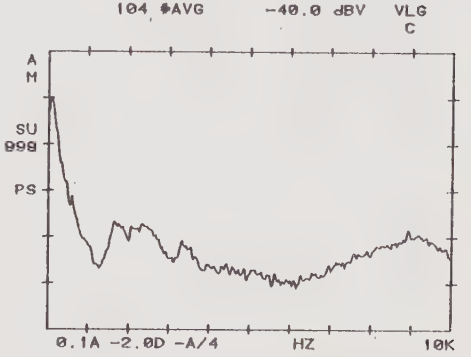


Abb. 3.10: Isoliert gesprochenes [f,], Sprecher 1, Realisation 1

mit den Rechnungen und Messungen von FANT (zumindest für [f,]) sowie STEVENS und HEINZ überein.

Dieser Formant ist in den folgenden Abbildungen nicht so deutlich zu erkennen, da dort dem Zweck der Untersuchung entsprechend der untere Frequenzbereich gespreizt wurde. Dadurch werden die meisten Darstellungen auf 8 kHz begrenzt. Die Abbildungen 3.8 und 3.10 zeigen den Bereich bis 10 kHz, während Abbildung 3.9 eine Spreizung bis 1000 Hz darstellt. Abbildung 3.8 zeigt das Spektrum für [f]. Es handelt sich hierbei um eine von Abbildung 3.7 verschiedene Realisation, also nicht etwa um eine Spreizung, wie aufgrund der guten Übereinstimmung der Formantpositionen leicht angenommen werden könnte. Diese gute Reproduzierbarkeit kann als Hinweis dafür aufgefaßt werden, daß die Analyse isoliert produzierter /f/-Laute bei sorgfältig eingestellter Analysetechnik und einiger Übung der Sprecher trotz der bekannten Probleme keineswegs von vornherein abzulehnen ist. Zwar läßt sich so wenig über die Eigenschaften des /f/ in fortlaufender Sprache ermitteln, aber mindestens zur meßtechnischen Verifikation akustisch-phonetischer Theoriebildungen lassen sich sicher wertvolle Hinweise ableiten.

Aus den Bildern 3.8 und 3.10, deren Formantstruktur praktisch identisch ist, lassen sich die in Tabelle 3.1 augeführten Werte ablesen:

Minimum (Hz)	Maximum (Hz)
1300	1700
2000	2400
3200	3500
4200	4600
6000	9200

Tabelle 3.1: Formantstruktur im isoliert gesprochenen /f/

Die deutlichen Abweichungen im Spektrum von [f] zu [f,], wie sie in FANTs Messungen zu sehen sind, sind hier nicht zu beobachten.

3.2.3 Vergleichende Gegenüberstellung verschiedener Messungen

In der Diskussion der Spektralstruktur der Frikative kann es eine einzelne, für das /f/ charakteristische Darstellung für alle Randbedingungen nicht geben. Wie an späterer Stelle gezeigt wird, ist der Einfluß des lautlichen Kontextes ausreichend, Formantpositionen erheblich zu verändern. Um dennoch eine Art Grundstruktur aufzeigen zu können sind in Abbildung 3.11 die Ergebnisse verschiedener Autoren sowie eigene Messungen einander gegenübergestellt. Aus der Arbeit von STEVENS und HEINZ wurde die Messung für /fear/ gewählt. Aus den eigenen Arbeiten wurden die Werte aus der Mittelung über alle Sprecher, Realisationen und Kontexte (Typ 5, siehe Abschnitt 2.4 bzw. Abb. 5.12) eingetragen, weil hier die geringsten, nicht vom /f/ stammenden Einflüsse zu erwarten sind (siehe Kapitel 5). Die Abbildung zeigt, daß zwischen 1 kHz und 6 kHz bei allen Autoren dieselbe Anzahl von Minima und Maxima zu finden ist, diese allerdings - besonders im höheren Frequenzbereich - deutlich gegeneinander verschoben sind. Eine nahezu identische Lage ergibt sich für die Maxima bei 1700 Hz und 2400 Hz und den Minima bei 1300 Hz und 1900 Hz. Die nicht mehr so gute Übereinstimmung bei höheren Frequenzwerten kann u.a. durch die Variabilität bei individuellen Realisationen erklärt werden. Aufgrund der großen Zahl der Realisationen, Kontexte und Sprecher, die der Mittelung des Typs 5 zugrunde liegt, dürfte diese Messung gegenüber den anderen vorliegenden Daten am besten die Spektralstruktur des /f/ wiedergeben.

Abb. 3.11: Spektralstruktur des /f/, Gegenüberstellung der Ergebnisse verschiedener Autoren

4. BERECHNUNG DER /f/-SPEKTREN

In diesem Kapitel werden die theoretischen Grundlagen für die Experimente dieser Arbeit bereitgestellt. Im ersten Teil wird kurz auf die mathematischen Grundlagen der Fouriertransformation eingegangen, insbesondere auf diejenigen Aspekte, die sich bei der Verarbeitung stochastischer Signale ergeben. Der zweite Teil befaßt sich mit den möglichen Aufbereitungsarten der Einzelergebnisse wie Fragen der Normierung, der Mittelungsarten usw.

4.1 Mathematische Grundlagen

4.1.1 Anmerkungen zur Fouriertransformation

Die Fouriertransformation hat in der Entwicklung der Signalverarbeitung aufgrund der ihr innewohnenden Eigenschaften eine namhafte Rolle gespielt und speziell für die Sprachverarbeitung außerordentliche Konsequenzen gehabt. Sie erlaubt unter bestimmten Umständen die ein-eindeutige Transformation eines Zeitsignals aus dem Zeitbereich in den sogenannten Frequenzbereich und liefert dabei u.a. eine Analyse der Frequenzkomponenten, aus denen sich das Signal zusammensetzt. Ihre besondere Eignung für die Sprachverarbeitung wird durch einen Blick auf das gebräuchliche nachrichtentechnische Modell der Sprachproduktion deutlich, das in Abbildung 4.1 skizziert ist.
Danach entsteht das Ausgangssignal durch Anregung des Vokaltraktsystems durch zwei Signalquellen, einer periodischen und einer Rauschquelle, die einzeln oder gemeinsam aktiv sein können. Der Vokaltrakt wirkt durch Resonanzeffekte als Filter und verstärkt bzw. dämpft dadurch die Energie in einzelnen Frequenzbereichen. Die Resonanzfrequenzen des Vokaltraktes sind im Signal als Formanten zu beobachten.

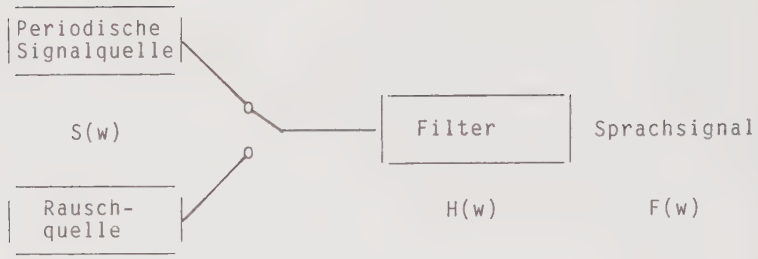


Abb. 4.1: Modell der Sprachlautproduktion

Phonetisch gesehen entspricht der periodischen Signalquelle die Phonation an den Stimmlippen, während die stochastische Quelle die Vorgänge an der Konstriktion bei Frikativen beschreibt. Bei einem solchen System ist das Spektrum des Ausgangssignals F(w) das Produkt von Quellenspektrum S(w) und der das Filter (den Vokaltrakt) beschreibenden Übertragungsfunktion H(w). Die mathematische Beschreibung des Modells in Abbildung 4.1 bietet damit die Möglichkeit, die Entstehung der Ausgangsgrößen aus den Eingangsdaten in einfacher Weise darzustellen. Der Übergang aus dem Zeit- in den Frequenzbereich wird durch eine Fouriertransformation vollzogen. Die Theorie der Fouriertransformation ist in der einschlägigen Literatur ausführlich und auch unter den verschiedensten Anwendungsaspekten dargestellt. Daher sollen an dieser Stelle nur die zum Verständnis dieser Arbeit wichtigsten Begriffe und Eigenschaften in Erinnerung gerufen werden.

Eine kontinuierliche periodische Zeitfunktion f(t) mit der Periodendauer T läßt sich in einer Reihenentwicklung darstellen (LEE 1960):

$$(1) \qquad f(t) = \sum_{k=-\infty}^{k=+\infty} F(k)\, e^{jkwt} \quad \text{mit } w = \frac{2\pi}{T}$$

bzw. $$f(t) = \sum_{k=-\infty}^{k=+\infty} |F(k)|\ e^{j(kwt+P(k))}$$

und

(2) $$F(k) = 1/T \int_{0}^{T} f(t)\ e^{-jkwt} dt$$

Die Fourierkoeffizienten F(k) sind die komplexen Amplituden des diskreten Spektrums von f(t). Auf diese Weise entsteht eine Frequenzanalyse, da die Koeffizienten F(k) die im Zeitsignal f(t) enthaltenen Frequenzanteile beschreiben. In der Darstellung (2) heißen |F(k)| Amplitudenspektrum und P(k) Phasenspektrum von f(t).

In der Praxis liegt nun aber kein periodisches Signal unendlicher, sondern nur endlicher Dauer vor, das zudem in der Regel nur in kurzen Abschnitten periodisch ist. Bei Verarbeitung mit Digitalrechnern wird das Signal außerdem in eine Folge zeitlich äquidistanter Amplitudenwerte $f_i=(t_0+i\Delta t)$ zerlegt und analog/digital umgesetzt. Es liegt also in einem solchen Fall kein kontinuierliches, sondern ein diskretes Signal vor. Hierfür gelten andere Betrachtungen als sie für kontinuierliche Signale gültig sind.

Die Zahlenfolge f(n) sei definiert in einem Bereich $0 \leq n \leq N$. Indem zunächst formal angenommen wird, daß sich die endliche Punktfolge f(n) von $-\infty$ bis $+\infty$ periodisch fortsetzt, lassen sich entsprechende Relationen ableiten, die als Diskrete Fourier-Transformation bezeichnet werden (DFT).

(3a) $$F(k) = \sum_{n=0}^{N-1} f(n)\ e^{-jwkn} \qquad 0 \leq k \leq N-1$$

$$\text{mit } w = \frac{2\pi}{N}$$

$$(3b) \quad f(n) = 1/N \sum_{k=0}^{n-1} F(k)\, e^{+jkwn} \qquad 0 \leq n \leq N-1$$

Bezüglich der Eigenschaften dieser DFT im einzelnen sei auf die diesbezüglich sehr umfangreiche Literatur verwiesen, z.B. auf die Darstellungen in OPPENHEIMER & SCHAFER (1975, Seite 110 ff).

Für die praktische Anwendung ist das Gleichungssystem (3) noch nicht geeignet, da sich in der physikalischen Wirklichkeit die Punktfolge f(n) nicht periodisch sondern verändert fortsetzt. Dem oben genannten Ansatz folgend wird die Vorausetzung für die Anwendbarkeit der DFT dadurch geschaffen, daß aus dem Signal ein geeignetes Segment herausgenommen und - gedanklich - nach links und rechts unendlich oft wiederholt wird. Dieses Segment sollte in seiner Länge einerseits groß genug sein physikalische Ereignisse zu repräsentieren, andererseits aber auch nicht zu groß sein, damit die im zeitlichen Verlauf auftretenden den physikalischen Ereignissen entsprechenden Signalveränderungen nicht durch Mittelung verschwinden. Diese Aufgabe übernimmt eine Fensterfunktion (Window), deren einfachste Form das Rechteckfenster

$$w(n) = 1 \quad \text{für} \quad 0 \leq n \leq N-1$$
$$w(n) = 0 \quad \text{sonst}$$

ist, das jeden Punkt im Signalsegment gleichermaßen gewichtet. Die an den Rändern entstehenden Sprungstellen führen allerdings zu einer Verfälschung des Spektrums. Aus diesem Grund zeichnen sich in der Regel die in der Praxis verwendeten Fenster durch flach ansteigende Flanken aus, womit eine unterschiedliche Wichtung der einzelnen Werte vorgenommen wird. Ein häufig verwendetes Fenster ist z.B. das "Hamming-Window":

$$w(n) = 0.54 - 0.46 \cos(2\pi n/(n-1)) \quad \text{für} \quad 0 \leq n \leq N-1$$
$$w(n) = 0 \quad \text{sonst}$$

Die cosinusförmige Gestalt dieses Fensters wichtet die mittleren Signalpunkte stärker als die an den Rändern gelegenen, so daß außerdem Sprünge von Segment zu Segment vermieden werden. Die weiteren Eigenschaften unterschiedlicher Fenster, die über die Selektion des Zeitsegmentes hinausgehen, sind in der Literatur ausführlich dargestellt (z.B. RABINER & SCHAFER 1978).
Für die diskrete Fouriertransformation ergibt sich also

$$(4) \qquad F(k) = 1/N \sum_{n=0}^{N-1} w(n)\, f(n)\, e^{-jwnk}$$

und für das reelle Leistungsdichtespektrum

$$(5) \qquad S(k) = F(k)^{*}F(k) = 1/N\, |F(k)|^{2}$$

wobei $F^{*}(k)$ die konjugiert komplexe Funktion von F(k) ist.

4.1.2 Anwendung der Fouriertransformation auf stochastische Signale

Die Uberlegungen des vorstehenden Abschnittes gelten für deterministische Signale und sind nicht ohne weiteres auf stochastische Prozesse anwendbar. Signale werden als stochastisch bezeichnet, wenn die betrachtete physikalische Meßgröße als Funktion der Zeit einen nicht vorhersehbaren - in der Literatur oft "regellos" genannten - Verlauf aufweist. Ein bestimmtes stochastisches Signal läßt sich auch in gleichartigen Experimenten nicht wiederholen. Eine geschlossene mathematische Formulierung, die den Verlauf des Signals beschreibt und damit als Grundlage für weitere Berechnungen dienen könnte, ist nicht angebbar. Folglich sind Betrachtungen eines einzelnen stochastischen Signals zunächst unergiebig. Wenn aber das einzelne derartige Signalsegment als Repräsentant (Musterfunktion) einer Klasse von unendlich vielen regellosen Signalen aufgefaßt wird,

die durch einen bestimmten Erzeugungsprozeß entstehen, wird die Bedeutung seines Spektrums zur Beschreibung seiner Klasse sinnvoll.

In diesem Abschnitt wird in groben Zügen gezeigt, welche Probleme auftreten, wenn die Leistungsdichtespektren solcher Signale berechnet werden müssen. Auch hier bietet die Literatur viele Darstellungen des Problems (OPPENHEIMER & SCHAFER 1978, LEE 1960, SCHLITT & DITTRICH 1972). Für die Berechnung der /f/-Spektren wird angenommen, daß das Zeitsignal als Musterfunktion eines stochastischen Prozesses angesehen werden kann. Diese Annahme ist auch dann zutreffend, wenn zusätzlich zu der das regellose Signal produzierenden Rauschquelle zeitweilig weitere periodische Signalquellen auftreten (siehe Kapitel 3), da bei stochastischen Anteilen das Zeitsignal nicht exakt vorhersehbar bleibt.

Wenn auch für das stochastische Signal keine mathematische Beschreibung im Zeitbereich angebbar ist, so lassen sich dennoch Kenngrößen angeben, wie im folgenden gezeigt werden soll.

f(t) sei eine Musterfunktion eines stochastischen Prozesses, definiert im Bereich von $t = -\infty$ bis $+\infty$. Dann heißt die Funktion

$$(6) \quad A(\tau) = \lim_{T \to \infty} 1/2T \int_{-T}^{+T} f(t)\, f(t+\tau)\, dt$$

Autokorrelationsfunktion von f(t). Ihr kommt bei der Berechnung des Leistungsdichtespektrums besondere Bedeutung zu, da sich nachweisen läßt, daß sie für alle Musterfunktionen desselben stochastischen Prozesses identisch ist (z.B. SCHLITT & DITTRICH 1972) und insofern als für den Prozeß repräsentativ betrachtet werden kann.

Es läßt sich nun zeigen (LEE 1960), daß zwischen dem Leistungsdichtespektrum S(w) des Signals f(t) und seiner Autokorrelationsfunktion $A(\tau)$ gilt

$$(7) \quad S(w) = 1/2\pi \int_{-\infty}^{+\infty} A(\tau)\, e^{-jw\tau}\, d\tau$$

Das Leistungsdichtespektrum eines stochastischen Signals f(t) ist also die Fouriertransformierte seiner Autokorrelationsfunktion $A(\tau)$. Im Gegensatz zur Analyse periodischer Signale ergibt sich kein diskretes Leistungsspektrum, sondern eine kontinuierliche Funktion, die die Leistungsdichte in Abhängigkeit von der Frequenz angibt.

Eine auf solche Weise durchgeführte Berechnung der Spektren bewirkt allerdings, daß die Phaseninformation nicht vorliegt und somit keine Rückschlüsse auf das komplexe Amplitudenspektrum möglich sind. Für die Praxis bei der Verarbeitung von Signaltypen wie das Sprachsignal, die zeitlich rasch aufeinanderfolgend periodische und stochastische Signalsegmente (und Mischformen) enthalten, ist es außerdem kaum durchführbar, ständig zwischen verschiedenen Berechnungsarten umzuschalten. Nun läßt sich aber zeigen, daß unter bestimmten Voraussetzungen für zeitlich begrenzte Signale die direkte Anwendung der Fouriertransformation auf das Zeitsignal als Abschätzung für das Leistungsdichtespektrums stochastischer Signalkomponenten zuläßig ist (SCHWARTZ & SHAW 1975).

Entsprechend Gleichung (7) gilt für diskrete Signale

$$(8) \quad S(w) = \sum_{k=-\infty}^{+\infty} A(k)\, e^{-jkwT}$$

wobei

$$(9) \qquad A(k) = 1/N \sum_{n=-\infty}^{+\infty} f(n)\, f(n+k)$$

die Autokorrelationsfunktion von der Folge f(n) ist. Wird f(n) durch ein Fenster w(n) zeitlich begrenzt, so daß

$$f^W(n) = \begin{cases} w(n)\, f(n) & \text{für } n = 0,1,....N-1 \\ 0 & \text{sonst} \end{cases}$$

folgt für die Autokorrelationsfunktion

$$(10) \qquad A(k) = 1/N \sum_{n=-\infty}^{+\infty} f^W(n)\, f^W(n+k)$$

Gleichung (10) eingesetzt in (8) ergibt

$$(11) \qquad S(w) = 1/N \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} f^W(n)\, f^W(n+k)\, e^{-jkwT}$$

Unter Benutzung der Identität $1 = e^{jwnT}\, e^{-jwnkT}$ wird Gleichung (11) zu

$$(12a) \quad S(w) = 1/N \sum_{n=-\infty}^{+\infty} f^W(n)\, e^{jwnT} \sum_{k=-\infty}^{+\infty} f^W(n+k)\, e^{-jw(n+k)T}$$

$$(12b) \quad S(w) = 1/N \sum_{n=-\infty}^{+\infty} w(n) f(n)\, e^{jwnT} \sum_{k=-\infty}^{+\infty} w(n+k) f(n+k)\, e^{-jw(n+k)T}$$

Durch Ändern der Summationsvariablen n + k = m ergibt sich mit Gleichung (4) aus Gleichung (12)

$$(13) \qquad S(w) = 1/N\ F(k)^{*}F(k) = 1/N\ |F(k)|^{2}$$

Der Vergleich mit Gleichung (5) zeigt, daß die Fouriertransformation in der geschilderten Weise auch bei stochastischen Signalen unmittelbar auf das Zeitsignal angewendet werden darf.

4.1.3 Die Fast-Fourier-Transformation

Die Fast-Fourier-Transformation (FFT) ist lediglich ein Mittel zur schnellen und effizienten Berechnung der DFT. Für eine Folge von N Signalpunkten müssen zur Ausführung einer DFT N**2 komplexe Multiplikationen und N(N-1) komplexe Additionen ausgeführt werden (OPPENHEIMER & SCHAFER 1975, Kapitel 6). Ausgehend von einem für viele Untersuchungszwecke geeigneten Sprachsegment von etwa 40 msec Dauer ergeben sich bei Digitalisierungsfrequenzen von z.B. 12.8 kHz und 25.6 kHz, die u.a. für die vorliegende Untersuchung zur Verfügung standen, die Werte N = 512 bzw. N = 1024. Daraus wird die hohe Zahl an Rechenoperationen für eine direkte Berechnung der DFT offensichtlich, die auch mit modernen Rechenanlagen nicht effizient oder gar in Echtzeit durchgeführt werden kann. Einen großen Fortschritt brachte ein von COOLEY & TURKEY (1965) veröffentlichter Algorithmus, der den Beginn der Entwicklung einer Vielzahl von Methoden zur schnellen Berechnung der DFT einleitete, die dann später als Fast-Fourier-Transforms (FFT) den praktischen Einsatz der Fouriertransformation ermöglichten. Voraussetzung war, daß die Zahl N in ein Produkt aus mindestens zwei natürlichen Zahlen zerlegbar ist; eine Bedingung, die in der Praxis leicht zu erfüllen ist. Die erreichbaren Zeitreduktionen sind je nach Algorithmus unterschiedlich und bewegen

sich zwischen einer direkten Proportionalität zwischen der Zahl der Rechenoperationen und N^2, aber mit einem geringeren Proportionalitätsfaktor, und einer Proportionalität zu etwa $N*\log(N)$.

Zwischen der Fouriertransformation kontinuierlicher Signale und der Diskreten-Fourier-Transformation bestehen also aufgrund des Signals beachtliche Unterschiede. Dagegen stellt die DFT "nur" ein geeignetes Mittel zur schnellen Berechnung der FFT dar. Eine ausführliche Darstellung der Algorithmen zur FFT ist allein schon aufgrund des Umfangs der zugrunde liegenden Theorie nicht möglich. Auch hier sei wieder auf die umfangreiche, bereits zitierte Fachliteratur verwiesen.

4.2 Vorgehensweise und exemplarische Ergebnisse

4.2.1 Einzelspektren

Ein wesentliches Ergebnis der modernen Experimentalphonetik ist die Erkenntnis, daß Sprachlaute als Segmente im Sprachkontinuum dem Einfluß vorausgehender und/oder folgender Laute unterworfen sind, teilweise über mehrere Laute hinweg ("Koartikulation", siehe z.B. MENZERATH und DE LACERDA 1933, KOZHEVNIKOV und CHISTOVICH 1966). Von bestimmten lautlichen Umgebungen wird ein systematischer Einfluß auf die /f/-Spektren ausgehen, aufgrund dessen eine Klassenbildung möglich wird. Für die akustischen Korrelate der verschiedenen Typen von /f/-Spektren wurde in Abschnitt 1.3 der Begriff "/f/-Laute" eingeführt und zwar unabhängig davon, ob ein unterschiedlicher Höreindruck wahrnehmbar ist oder nicht. Für die Zwecke dieser Arbeit ist der kontextuelle Einfluß unerwünscht. Folglich ist eine möglichst gute Segmentierbarkeit und eine sorgfältige Durchführung der Segmentation von außerordentlicher Bedeutung.

Kontextuelle Einflüsse werden in der Regel zu Beginn und am Ende des /f/-Segmentes stärker sein als in der Mitte. Diese Überlegung wird durch die in Kapitel 3 beschriebenen Messungen von FANT bestätigt. Deshalb wurden die Programme zur Berechnung der Spektren in einer Version so gestaltet, daß nach der Segmentation des /f/ zur Berechnung der Spektren automatisch ein Signalsegment von 20 msec Dauer aus der Mitte einer jeden /f/-Realisierung gewählt wurde. Die stärker kontextuell beeinflußten Ränder des /f/-Zeitsignals werden auf diese Weise nicht in die Berechnung der Spektren einbezogen. Es ergibt sich für jedes /f/ ein einzelnes Spektrum, das im folgenden als "Einzelspektrum" bezeichnet wird.
Um einen Eindruck dieses Spektrumtyps zu vermitteln, zeigen die Abbildungen 4.2 (Vater) und 4.3 (schläft) als Beispiele zwei Einzelspektren. Weitere Einzelspektren für verschiedene Kontexte mit einer Darstellung des dazugehörigen Zeitsignals finden sich in den Abbildungen 4.10 bis 4.15. In den Abbildungen 4.16

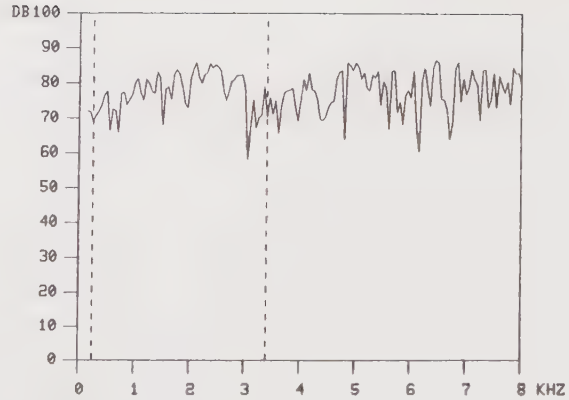


Abb. 4.2: Einzelspektrum, Sprecher 1, Kontext 1 (Vater), Realisation 1

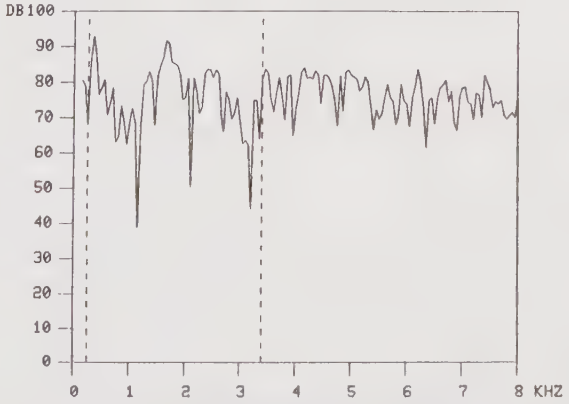


Abb. 4.3: Einzelspektrum, Sprecher 1, Kontext 2 (schläft viel), Realisation 1

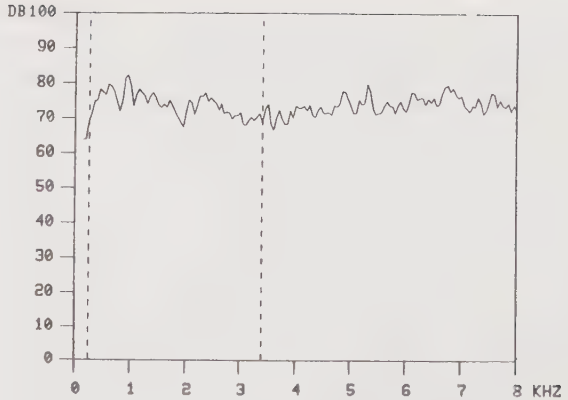


Abb. 4.4: Mittelwertspektrum, Sprecher 1, Kontext 1 (Vater), Realisation 1

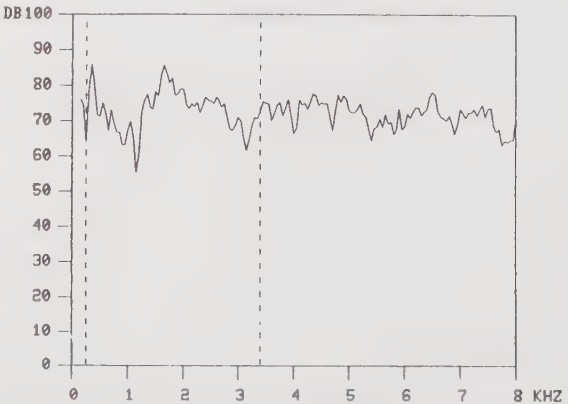


Abb. 4.5: Mittelwertspektrum, Sprecher 1 Kontext 2 (schläft viel), Realisation 1

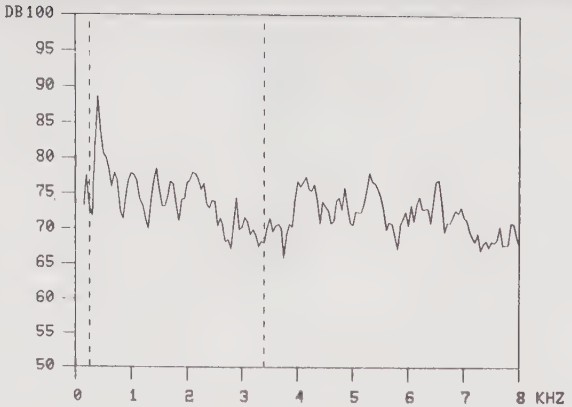


Abb. 4.6: Mittelwertspektrum, Sprecher 1, Kontext 4 (vielfach), Realisation 1

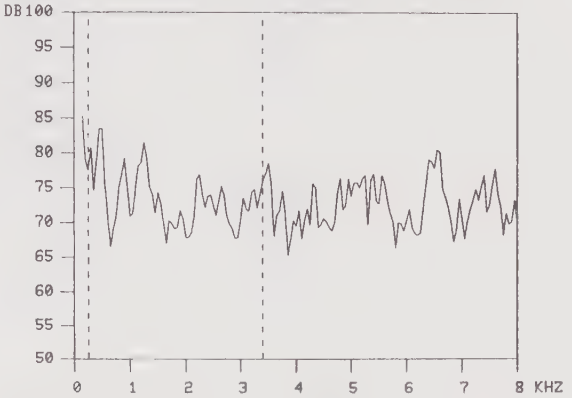


Abb. 4.7: Mittelwertspektrum, Sprecher 2, Kontext 4 (vielfach), Realisation 1

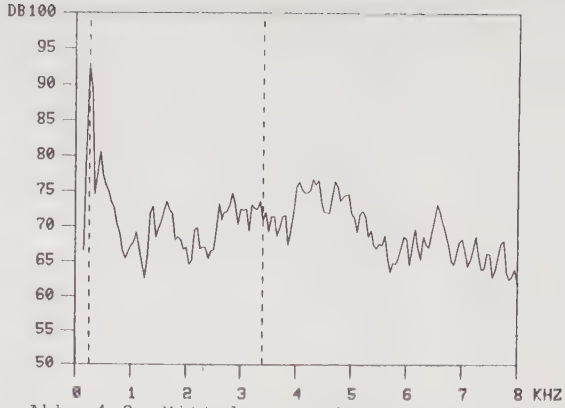


Abb. 4.8: Mittelwertspektrum, Sprecher 3, Kontext 4 (vielfach), Realisation 1

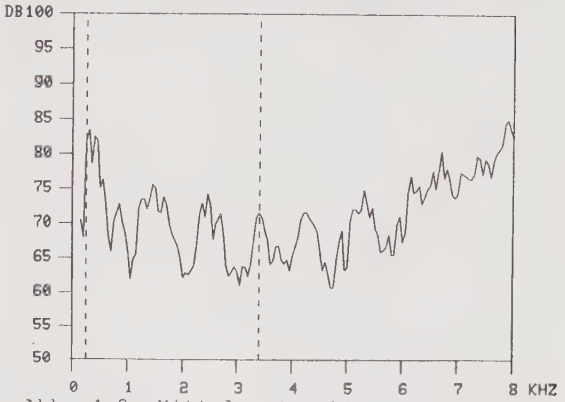


Abb. 4.9: Mittelwertspektrum, Sprecher 4 Kontext 4 (vielfach), Realisation 1

10 dB/Skt

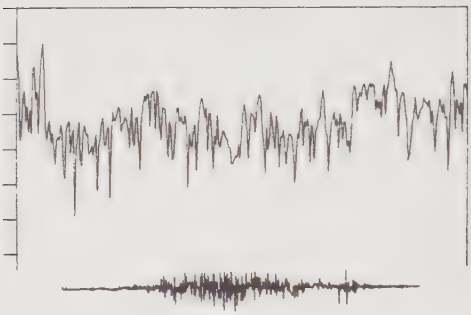


Abb. 4.10:
Einzelspektrum
Sprecher WOE
Kontext 3: schläft viel
Dauer 74 msec

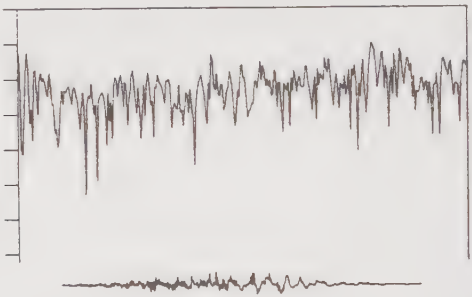


Abb. 4.11:
Einzelspektrum
Sprecher WOE
Kontext 4: vielfach
Dauer 31 msec

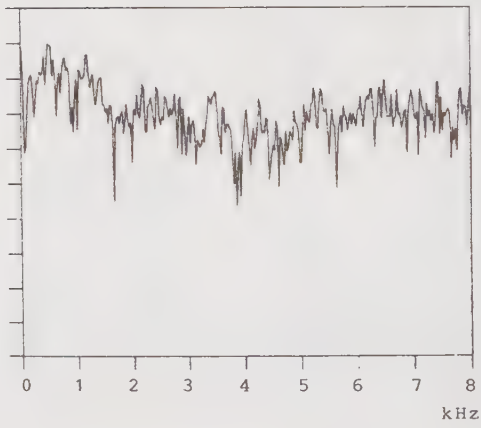


Abb. 4.12:
Einzelspektrum
Sprecher WOE
Kontext 5: fach vor
Dauer 49 msec

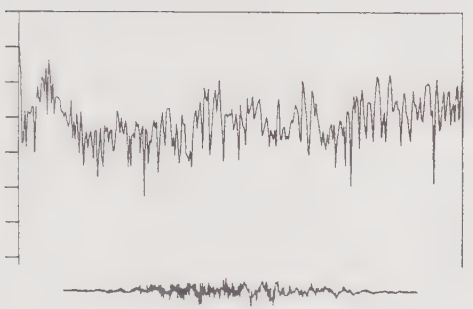


Abb. 4.13:
Einzelspektrum
Sprecher WOE
Kontext 7: das fällt
Dauer 75 msec

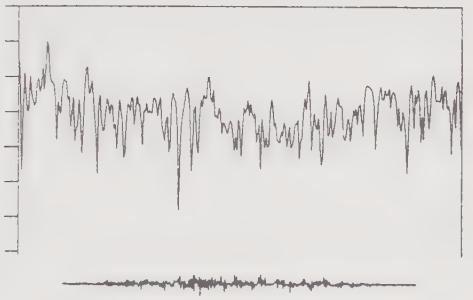


Abb. 4.14:
Einzelspektrum
Sprecher WOE
Kontext 8: der Familie
Dauer 51 msec

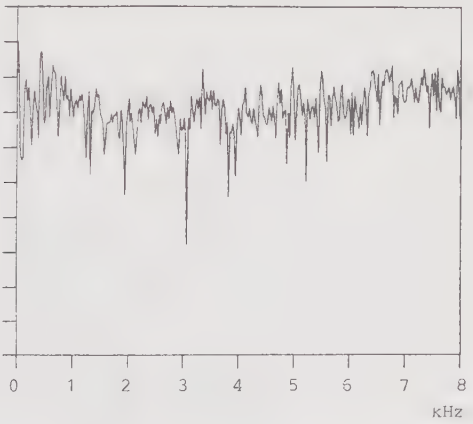


Abb. 4.15:
Einzelspektrum
Sprecher WOE
Kontext 9: auf
Dauer 221 msec

und 4.17 sind Einzelspektren den im nächsten Abschnitt beschriebenen Mittelwertspektren gegenübergestellt. Es handelt sich dabei um zwei Realisationen des /f/ in Kontext 2 (schläft viel) desselben Sprechers. Eine Diskussion der Spektren erfolgt in Kapitel 5.

4.2.2 Mittelwertspektren

Für die spätere Anwendung der Ergebnisse dieser Untersuchung ist das Intervall von großer Bedeutung, in dem sich die Amplituden der einzelnen Spektren an einem bestimmten Frequenzwert bewegen. Dieses Intervall wird in dieser Arbeit als Streubreite bezeichnet. Da nun immer wieder auftretende "Spitzen" im Spektrum die Streubreite entscheidend bestimmen, kann alternativ zu der im vorhergehenden Abschnitt beschriebenen Vorgehensweise für jede /f/-Realisierung ein Mittelwertspektrum aus sich überlappenden Lautanteilen gebildet werden. Dazu wird das Hamming-Fenster mit 1/3 Überlappung von Beginn bis Ende über das Signalsegment geschoben und eine Mittelung der sich so ergebenden Einzelspektren durchgeführt (siehe hierzu auch die Ausführungen in GOYDKE 1973). Auf diese Weise wird unter eventueller Inkaufnahme größeren kontextuellen Einflusses eine erwünschte glättende Wirkung auf den Verlauf der Spektren erzielt. Die Abbildungen 4.4 und 4.5 zeigen Mittelwertspektren für dieselben /f/-Realisationen wie in den Abbildungen 4.2 und 4.3. Der erheblich glattere Verlauf der Spektren wird im Vergleich deutlich. Vier weitere Mittelwertspektren sind in den Abbildungen 4.6 bis 4.9 dargestellt. Es handelt sich dabei jeweils um das /f/ in Vater aus der ersten Realisation der Sprecher 1 bis 4. Auf die Gegenüberstellung mit Einzelspektren in den Abbildungen 4.16 und 4.17 sei ebenfalls hingewiesen.

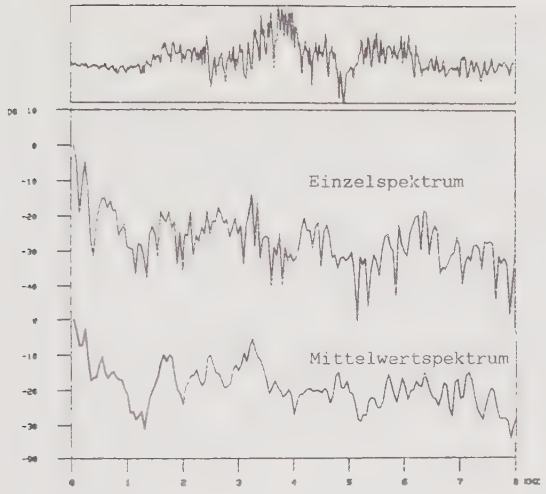


Abb. 4.16: Einzel- und Mittelwertspektrum, Kontext 2 (schläft viel), Sprecher 1, Realisation 1, Dauer 51 msec

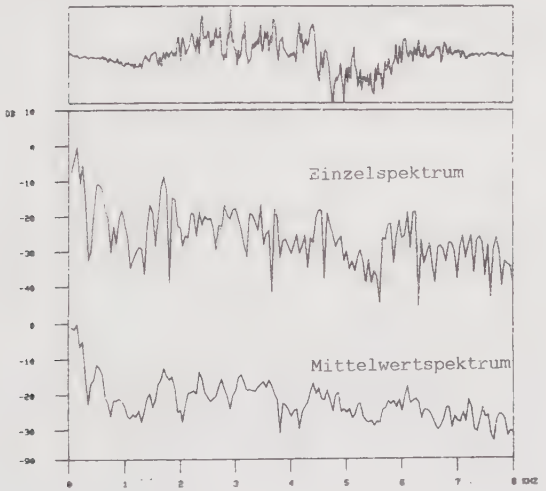


Abb. 4.17: Einzel- und Mittelwertspektrum, Kontext 2 (schläft viel), Sprecher 1, Realisation 2, Dauer 50 msec

4.2.3 Normierung der Spektren

Für den Vergleich untereinander und vor allem für die Ermittlung der Streubreiten ist eine Normierung der Spektren unerläßlich. Ein mögliches Verfahren ist die Normierung auf einen einheitlichen Mittelwert, wie sie für sämtliche vorliegenden Spektren durchgeführt wurde. Da es für die vorliegenden Ziele aber besonders auf den Telefonkanal zwischen 300 Hz und 3400 Hz ankommt, wurde in einer Voruntersuchung ermittelt, ob sich die Streuungen der Spektren relativ zueinander innerhalb des Telefonkanals verringern, wenn die Normierung ebenfalls nur diesen oder einen ähnlich eingeschränkten Bereich umfaßt. Es zeigte sich, daß bessere Werte erzielt werden, wenn der Bereich bis 200 Hz wegen der dort auftretenden Störeffekte, z.B. das Auftreten der 5o Hz Netzfrequenz und andere niederfrequenter Ereignisse, aus der Normierung herausgenommen wird. Dagegen verschlechtern sich die Streubreiten nur minimal, wenn statt 3.4 kHz obere Eckfrequenz 8.0 kHz gewählt werden. Bei einheitlicher Normierung von 150 Hz bis 8 kHz wurden die Streubreiten hierfür und getrennt davon für den Telefonkanal ermittelt.

4.2.4 Art der Mittelung

Wie vorstehend beschrieben, werden für jeden /f/-Laut Einzel- und Mittelwertspektren betrachtet. Die Darstellung und Diskussion der Ergebnisse erfolgt in der Einheit "dB". Da die Absolutwerte der Spektren innerhalb eines Frequenzkanals außerdem eine schiefe Verteilungsform bilden, die nach der für die Umrechnung in "dB" erforderlichen Logarithmierung eher einer Normalverteilung ähnelt und damit leichter zu untersuchen ist, wurden sämtliche normierten Spektren vor jeder weiteren Verarbeitung logarithmiert. Bei der späteren Anwendung muß ebenso verfahren werden. Die alternative Vorgehensweise, nämlich ausschließlich mit nichtlogarithmierten Werten zu arbeiten und diese nur zu Darstellungszwecken zu logarithmieren, wurde exemplarisch eben-

falls untersucht und führt erwartungsgemäß zu gleichwertigen Ergebnissen.

4.2.5 Auswahl des Datenmaterials und der Berechnungsmethode

Ein Kriterium bei der Auswahl der möglichen Kontexte kann neben phonetischen Gesichtspunkten die Häufigkeit des Auftretens in normaler Rede sein. Damit steigt nämlich die Wahrscheinlichkeit, daß sich bei der späteren Anwendung der Methode die Lautkombination aus der Tataufnahme im Hintergrundmaterial wiederfindet. Als Sprachmaterial für die Untersuchung wurde der Satz

"Vater schläft vielfach vor dem Fenseher ein,
das fällt der Familie auf."

verwendet. Damit konnten folgende Kontextkonfigurationen bzw. /f/-Laute untersucht werden:

Kontext 1: Vater
Kontext 2: schläft viel
Kontext 3: schläft viel
Kontext 4: vielfach
Kontext 5: fach vor
Kontext 6: dem Fernseher
Kontext 7: das fällt
Kontext 8: der Familie
Kontext 9: auf

Insgesamt wurden 10 Sprecher aufgenommen mit je 10 Realisationen in je einer Sitzung. Davon wurden fünf Sprecher auf Revox B-77 in einem reflexionsarmen Raum und fünf Sprecher auf Uher Report in verschiedenen Büroräumen aufgezeichnet. Hierfür waren ausschließlich praktische Gründe maßgebend; das Material wurde nicht in dieser Hinsicht untersucht. Die insgesamt neun /f/-Laute in den verschiedenen Kontexten ergeben bei 10 Sprechern und je

10 Realisationen 900 /f/-Realisierungen für die Untersuchungen. Abbildung 4.10 zeigt ein Sonagramm des Testsatzes.

Das Sprachmaterial für die Anwendungsuntersuchungen in Kapitel 6 wurde über verschiedene Orts- und Fernleitungen von privaten und von öffentlichen Telefonzellen aus übertragen. Die Sprachproben wurden teilweise direkt gesprochen, teilweise vom Band eingespielt. Bei den Sprechern handelte es sich auch um Personen, die nicht bei der Gewinnung des Hintergrundmaterials beteiligt waren. Eine nähere Beschreibung wird in Kapitel 6 im Zusammenhang mit den einzelnen Untersuchungen gegeben.

Zu jedem der 900 /f/-Laute wurden die dazugehörigen Spektren mittels FFT berechnet. Die Segmentation der /f/-Laute wurde an einem graphischen Bildschirm durchgeführt und die Segmentgrenzen in Dateien abgelegt, auf die die Analyseprogramme automatisch zugreifen können (siehe hierzu auch Abschnitt 5.3.1). Im Hinblick auf die Fragestellung wurden Voruntersuchungen bezüglich der Auflösung im Frequenzbereich bei der Berechnung der Spektren durchgeführt. Wird sie zu grob gewählt, werden nicht alle relevanten Ereignisse erfaßt; bei einer zu feinen Auflösung ergeben sich scharfe Linien, die später ungerechtfertigt hohe Streubreiten in den einzelnen Frequenzkanälen vortäuschen. Aufgrund visueller Vergleiche von Spektren verschiedener Auflösung wurden 50 Hz als optimal geeignet ermittelt, um Spektralstrukturen der /f/-Laute gut wiedergeben zu können. Für die FFT bzw. für die Analog/digital-Wandlung wurden folgende Parameter gewählt:

Digitalisierungsfrequenz:	25.6 kHz
Quantisierung:	12 bit
Filter:	10.0 kHz
Länge der FFT:	512 Werte
Fenster:	Hamming

Da die Untersuchungen nur bis 8.0 kHz geführt wurden, hätte grundsätzlich eine geringere Digitalisierungsfrequenz ausgereicht. Die verwendeten Frequenzen von 25.6 kHz bzw. 10.0 kHz

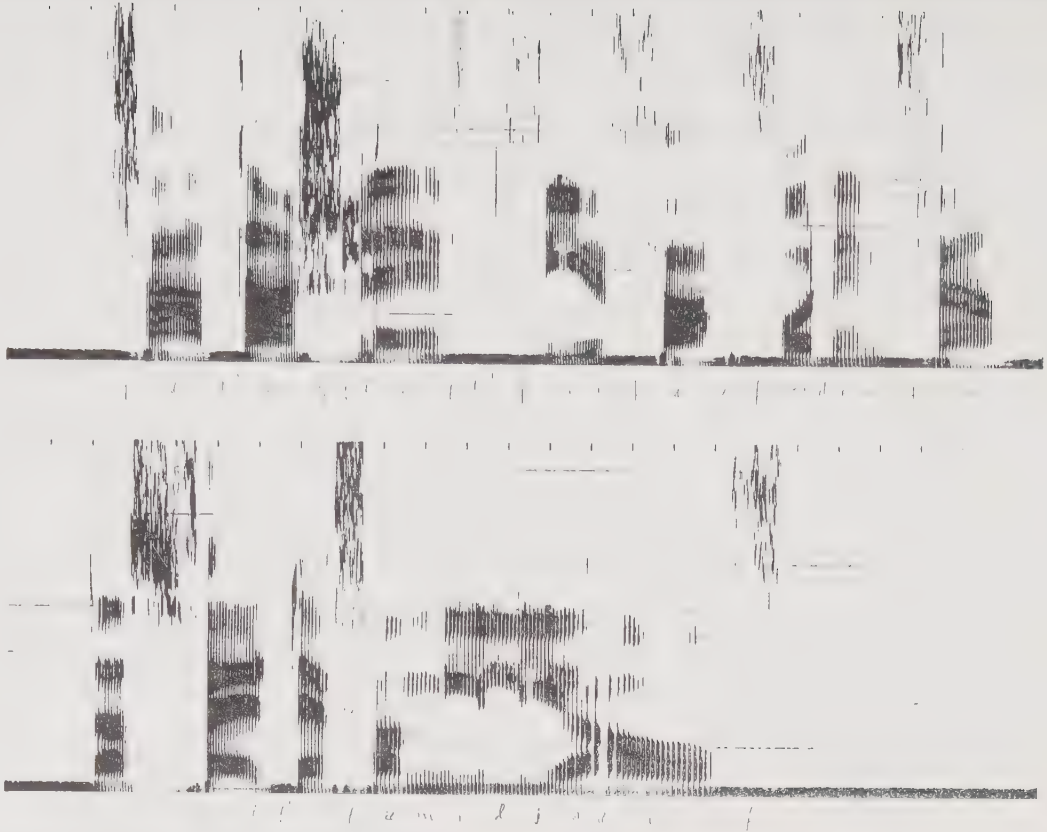


Abb. 4.18: Sonagramm einer Realisierung des Testsatzes "Vater schläft vielfach vor dem Fernseher ein, das fällt der Familie auf."

für das Filter ergaben sich aus praktischen Gründen aufgrund der vorhandenen Hardwarekonfiguration.

5. ERGEBNISSE

In diesem Kapitel werden die wesentlichen Meßergebnisse der experimentellen Untersuchung vorgestellt. Schwerpunkt sind die in Abschnitt 2.4 aufgezeigten Fragestellungen, also die Untersuchung der Spektralstruktur der /f/-Laute im Hinblick auf ihre Sprecher-, Kontext und Realisationsabhängigkeit. Vorangestellt wird eine kurze Betrachtung der Dauer der einzelnen /f/-Laute mit tabellarischer und graphischer Darstellung der Ergebnisse.

5.1 Dauer von /f/-Lauten

Für eine spätere Beurteilung der Zuverläßigkeit bei der Anwendung ist die Dauer der /f/-Laute bedeutsam, weil die Zahl der möglichen Mittelungen bei den Mittelwertspektren und damit der Kurvenverlauf u.a. davon abhängt. Das vorliegende Material gibt außerdem die Möglichkeit, die Abhängigkeit der Dauern von Kontexten, unterschiedlichen Sprechern und Realisationen zu untersuchen. In dieser Arbeit braucht der Frage im einzelnen nicht nachgegangen zu werden, da sich eine Berücksichtigung der Dauer bei der Bestimmung der Zuverlässigkeit der beschriebenen Methode in Anwendungsfällen indirekt durch die Abhängigkeit der Streubreiten von der Länge der /f/-Laute ergibt. Sollten nämlich bei kürzeren Segmenten die Streubreiten zu groß werden (soweit dies nicht durch andere Effekte, wie z.B. die Segmentumgebung, ausgeglichen wird) werden die zugrundeliegenden /f/-Realisierungen als wenig geeignet eingestuft und möglichst nicht verwendet. Im folgenden sollen daher nur in kurzer tabellarischer und graphischer Form die wichtigsten Meßergebnisse dargestellt werden.

Der Tabelle 5.1 können die Dauern aller vermessenen /f/-Realisierungen sowie die Mittelwerte (arithmetisch, ohne nähere Betrachtung der Verteilungsform) für die einzelnen Sprecher, innerhalb der einzelnen Kontexte, usw., entnommen werden. Der

NAME DER DATEI	KONTEXT DAUER DER SEGMENTE IN MSEC 1	2	3	4	5	6	7	8	9	MW
DB0:SAU01.SAM	97	56	68	125	115	75	59	88	195	97.6
DB0:SAU02.SAM	81	48	76	102	120	80	75	77	128	87.4
DB0:SAU03.SAM	118	53	73	94	111	68	74	59	149	88.8
DB0:SAU04.SAM	63	72	50	103	117	74	66	72	151	85.3
DB0:SAU05.SAM	94	56	75	85	87	60	45	80	183	85.0
DB0:SAU06.SAM	90	63	68	87	72	66	70	71	136	80.3
DB0:SAU07.SAM	118	54	73	116	110	65	66	59	125	87.3
DB0:SAU08.SAM	48	56	43	93	101	53	64	80	134	74.7
DB0:SAU09.SAM	90	79	85	100	89	53	63	49	143	83.4
DB0:SAU10.SAM	60	57	60	92	80	62	73	96	115	77.2
MITTELWERT	85.9	59.4	67.1	99.7	100.2	65.6	65.5	73.1	145.9	84.7
DB0:VAN01.SAM	30	51	92	80	53	53	54	25	218	72.9
DB0:VAN02.SAM	36	50	71	72	65	84	66	35	225	78.2
DB0:VAN03.SAM	42	50	80	81	147	53	65	54	150	80.2
DB0:VAN04.SAM	34	42	93	84	57	85	62	41	152	72.2
DB0:VAN05.SAM	39	56	94	89	62	63	39	61	190	77.0
DB0:VAN06.SAM	36	51	76	88	77	67	59	42	182	75.3
DB0:VAN07.SAM	35	57	81	81	149	75	56	39	187	84.4
DB0:VAN08.SAM	34	53	101	84	51	88	59	54	158	75.8
DB0:VAN09.SAM	20	41	78	102	62	69	60	69	199	77.8
DB0:VAN10.SAM	50	43	83	101	64	71	68	54	151	76.1
MITTELWERT	35.6	49.4	84.9	86.2	78.7	70.8	58.8	47.4	181.2	77.0
DB0:NZL01.SAM	122	75	137	101	76	91	91	96	139	103.1
DB0:NZL02.SAM	114	55	116	106	81	86	86	91	129	96.0
DB0:NZL03.SAM	106	83	167	108	108	91	89	97	118	107.4
DB0:NZL04.SAM	113	57	131	95	78	85	64	98	110	92.3
DB0:NZL05.SAM	118	50	103	110	83	101	86	100	117	96.4
DB0:NZL06.SAM	54	44	93	97	88	79	102	105	124	87.3
DB0:NZL07.SAM	115	63	131	103	94	83	83	91	140	100.3
DB0:NZL08.SAM	98	54	91	94	89	87	78	96	141	92.0
DB0:NZL09.SAM	75	38	89	99	77	108	82	104	106	86.4
DB0:NZL10.SAM	85	84	95	94	102	100	72	70	127	92.1
MITTELWERT	100.0	60.3	115.3	100.7	87.6	91.1	83.3	94.8	125.1	95.4
DB0:SLM01.SAM	35	37	118	96	65	37	42	29	164	69.2
DB0:SLM02.SAM	50	41	101	68	62	27	57	42	139	65.2
DB0:SLM03.SAM	33	42	104	71	73	38	48	51	121	64.6
DB0:SLM04.SAM	71	42	101	60	79	41	44	54	170	73.6
DB0:SLM05.SAM	59	41	109	60	62	31	64	37	169	70.2
DB0:SLM06.SAM	66	32	101	39	62	52	53	53	184	71.3
DB0:SLM07.SAM	73	29	103	66	79	47	61	56	166	75.6
DB0:SLM08.SAM	28	20	78	30	69	45	60	57	157	62.7
DB0:SLM09.SAM	81	26	101	75	80	52	64	50	127	72.9
DB0:SLM10.SAM	27	38	95	57	55	45	66	54	115	61.3
MITTELWERT	52.3	34.8	103.1	62.2	68.6	41.5	55.9	48.3	151.2	68.7

Tabelle 5.1: Dauern der /f/-Realisierungen
MW = Mittelwert der jeweiligen Zeile

NAME DER DATEI	DAUER DER SEGMENTE IN MSEC									
	1	2	3	4	5	6	7	8	9	MW
DB0:NIC01.SAM	101	42	108	104	71	78	75	55	122	84.0
DB0:NIC02.SAM	40	71	97	54	54	73	94	68	151	78.0
DB0:NIC03.SAM	109	81	75	82	156	83	71	79	168	100.4
DB0:NIC04.SAM	77	97	97	76	42	73	77	86	118	82.6
DB0:NIC05.SAM	109	79	119	97	51	68	73	59	99	83.8
DB0:NIC06.SAM	102	52	118	90	69	80	73	71	220	97.2
DB0:NIC07.SAM	67	57	105	102	34	85	50	93	131	80.4
DB0:NIC08.SAM	97	55	89	97	71	86	77	86	108	85.1
DB0:NIC09.SAM	86	57	89	71	123	90	102	63	141	91.3
DB0:NIC10.SAM	106	63	48	92	55	58	70	86	180	84.2
MITTELWERT	89.4	63.4	94.4	86.5	72.5	77.4	76.2	74.5	145.3	86.7
DB0:WOE01.SAM	129	80	53	58	90	98	79	96	105	87.6
DB0:WOE02.SAM	112	86	74	31	49	105	75	51	221	89.3
DB0:WOE03.SAM	76	71	149	105	90	100	101	84	118	99.3
DB0:WOE04.SAM	92	58	55	66	65	88	60	71	85	71.1
DB0:WOE05.SAM	99	123	88	98	98	114	65	89	143	101.9
DB0:WOE06.SAM	119	114	100	93	114	107	55	67	132	100.1
DB0:WOE07.SAM	125	91	98	75	97	105	83	58	170	100.2
DB0:WOE08.SAM	82	123	74	88	103	113	63	72	151	96.6
DB0:WOE09.SAM	95	115	109	69	63	89	67	91	175	97.0
DB0:WOE10.SAM	68	84	96	75	82	127	65	88	151	92.9
MITTELWERT	99.7	94.5	89.6	75.8	85.1	104.6	71.3	76.7	145.1	93.6
DB0:HEP01.SAM	69	60	91	100	151	49	49	87	211	96.3
DB0:HEP02.SAM	49	86	71	102	67	67	87	92	158	86.6
DB0:HEP03.SAM	84	70	84	91	89	81	68	75	122	84.9
DB0:HEP04.SAM	66	65	80	75	107	75	64	60	178	85.6
DB0:HEP05.SAM	65	66	86	76	58	50	60	88	136	76.1
DB0:HEP06.SAM	69	69	88	80	89	64	72	59	143	81.4
DB0:HEP07.SAM	86	96	82	81	94	83	53	72	149	88.4
DB0:HEP08.SAM	100	74	70	78	92	67	70	78	156	87.2
DB0:HEP09.SAM	59	80	75	91	91	85	74	74	136	85.0
DB0:HEP10.SAM	73	73	59	78	97	83	77	90	156	87.3
MITTELWERT	72.0	73.9	78.6	85.2	93.5	70.4	67.4	77.5	154.5	85.9
DB0:DEI01.SAM	85	79	47	78	97	61	38	39	190	79.3
DB0:DEI02.SAM	73	60	72	87	66	43	53	44	197	77.2
DB0:DEI03.SAM	74	70	80	102	71	50	56	54	162	79.9
DB0:DEI04.SAM	81	57	90	94	120	62	84	76	212	97.3
DB0:DEI05.SAM	73	55	82	80	70	59	48	49	217	81.4
DB0:DEI06.SAM	98	27	82	88	105	65	101	54	178	88.7
DB0:DEI07.SAM	70	44	92	84	81	67	55	52	234	86.6
DB0:DEI08.SAM	93	85	81	77	95	37	69	47	187	85.7
DB0:DEI09.SAM	66	39	115	86	93	77	67	48	178	85.4
DB0:DEI10.SAM	96	31	90	81	82	56	65	32	187	80.0
MITTELWERT	80.9	54.7	87.1	85.7	98.0	57.7	67.6	49.5	174.2	84.

Tabelle 5.1: Dauern von /f/-Realisierungen (Fortsetzung)

NAME DER DATEI	K O N T E X T									
	DAUER DER SEGMENTE IN MSEC									
	1	2	3	4	5	6	7	8	9	MW
DBO:PBE01.SAM	82	50	136	99	65	77	83	51	161	89.3
DBO:PBE02.SAM	90	67	119	89	81	55	91	70	191	94.8
DBO:PBE03.SAM	103	66	129	95	130	85	69	77	184	104.2
DBO:PBE04.SAM	87	64	142	100	92	83	83	90	187	103.1
DBO:PBE05.SAM	104	59	148	99	93	97	93	68	162	102.6
DBO:PBE06.SAM	94	72	149	104	66	86	91	60	178	100.0
DBO:PBE07.SAM	20	65	124	104	68	99	85	83	194	93.6
DBO:PBE08.SAM	98	65	126	93	67	83	90	97	138	95.2
DBO:PBE09.SAM	55	62	143	87	65	76	80	79	181	92.0
DBO:PBE10.SAM	59	53	119	90	74	91	93	83	158	91.1
MITTELWERT	79.2	62.3	133.5	96.0	80.1	83.2	85.8	75.8	173.4	96.6
DBO:LKL01.SAM	93	36	132	105	87	76	64	73	175	93.4
DBO:LKL02.SAM	99	34	81	98	103	64	65	57	171	85.8
DBO:LKL03.SAM	88	30	115	93	77	80	65	72	126	82.9
DBO:LKL04.SAM	85	38	114	98	89	70	60	73	156	87.0
DBO:LKL05.SAM	112	49	70	104	130	70	69	68	143	90.6
DBO:LKL06.SAM	91	46	118	85	69	70	51	80	145	83.9
DBO:LKL07.SAM	87	32	86	97	62	80	50	79	131	78.2
DBO:LKL08.SAM	75	23	91	86	61	75	59	68	133	74.6
DBO:LKL09.SAM	93	28	112	87	47	82	64	61	158	81.3
DBO:LKL10.SAM	80	28	139	93	96	66	40	64	139	82.8
MITTELWERT	90.3	34.4	105.8	94.6	82.1	73.3	58.7	69.5	147.7	84.0
MITTELWERT GESAMT	78.5	58.9	95.6	87.3	83.7	73.6	68.7	68.7	156.2	78.5

Tabelle 5.1: Dauern von /f/-Realisierungen
(letzter Teil)

Mittelwert über alle Sprecher, alle Realisationen und Kontexte beträgt 78.5 msec, entsprechend 10.3 Mittelungen der Spektren innerhalb eines Segmentes (bei den hier gewählten Randbedingungen). Dieser Wert stimmt mit der Angabe von GOYDKE (1973) überein, der die mittlere Dauer von /f/-Realisierungen zu etwa 80 msec bestimmte. Abbildung 5.1 zeigt die mittleren Segmentdauern der einzelnen /f/-Realisierungen in den verschiedenen Kontexten.

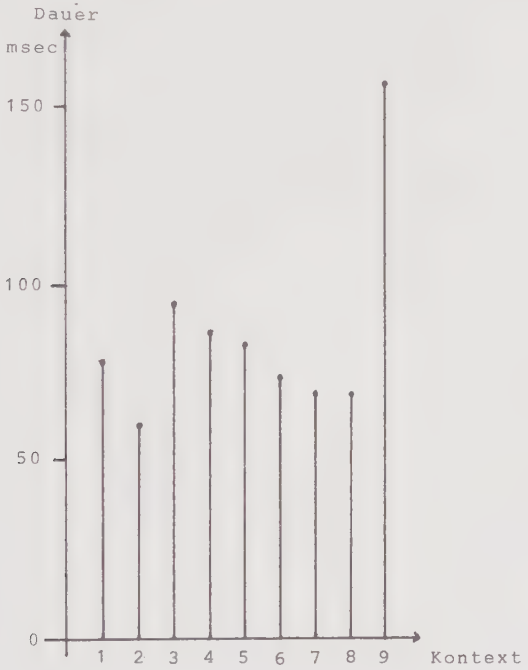


Abb. 5.1: Mittlere Dauern der /f/-Realisierungen in den verschiedenen Kontexten

Abbildung 5.2 verdeutlicht die sprecherspezifische Verteilung der Zeitdauern. Die Dauern einzelner Realisierungen einzelner Sprecher sind beispielhaft für den gut segmentierbaren Kontext 4 (vielfach) in Abbildung 5.3 dargestellt.

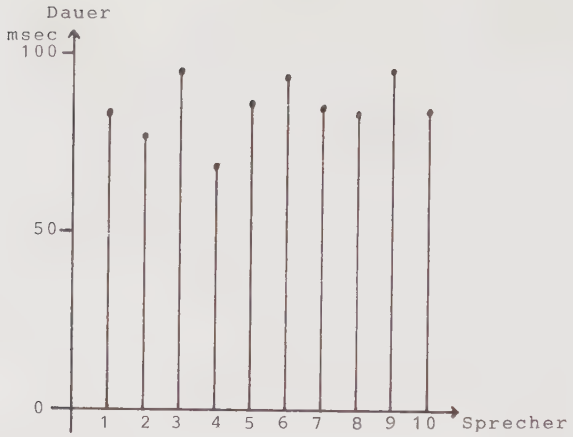


Abb. 5.2: Segmentdauern, gemittelt über alle Kontexte und Realisationen

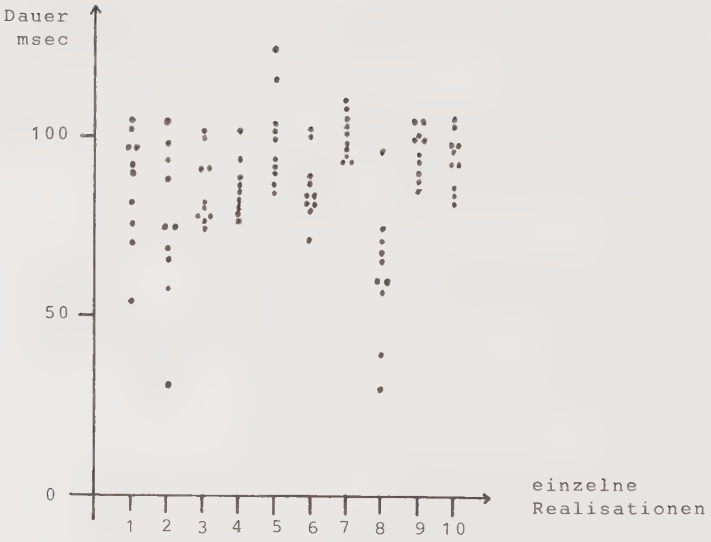


Abb. 5.3: Segmentdauern einzelner Sprecher und Realisationen für den Kontext 4

5.2 Spektren der /f/-Laute

5.2.1 Gegenüberstellung der Spektren

Die Spektren der /f/-Realisierungen wurden zunächst im Hinblick auf ihre akustische Struktur unter Berücksichtigung der Abhängigkeiten von unterschiedlichen Kontexten, Sprechern, etc. untersucht. In Abbildung 5.4 sind sämtliche Spektren der Mittelungstypen 1 bis 4 sowohl für Einzel- wie für Mittelwertspektren dreidimensional dargestellt. Dieses Bild gibt einen Überblick, wie sich Spektren von einer Mittelungsart zu einer anderen in ihrer Form verändern. Der Mittelungstyp 5, bestehend aus einer Kurve für alle Sprecher, Realisationen und Kontexte, wird später gesondert behandelt. Sämtliche Spektren umfassen den Frequenzbereich von 125 Hz bis 8.0 kHz; die Amplituden sind in dB geeicht. Zu jedem Bild sind der Minimal- und der Maximalwert der Amplituden in dB sowie die Differenz zwischen beiden angegeben.

Anhand der Abbildungen wird deutlich, daß Spektren vom Typ "Mittelwertspektren" eher als "Einzelspektren" geeignet sind, akustische Strukturen aufzuzeigen. Aufgrund ihrer erheblich größeren Zahl an Mittelungen werden "gröbere" Strukturen nicht durch Einzellinien verdeckt. Andererseits sind alle Ereignisse in beiden Formen wiederzufinden, wie z.B. das durch alle Spektren verlaufende Minimum bei etwa 5.5 kHz. Ausführliche Erläuterungen zu den einzelnen Abbildungen werden in den folgenden Abschnitten gegeben, die auch auf Abbildung 5.4 Bezug nehmen. Da die Werte der Mittelwertspektren naturgemäß weniger stark um ihren Mittelwert streuen, ist dieser Berechnungstyp vermutlich für allgemeine Aussagen über das /f/ sowie für die Zwecke dieser Arbeit besser geeignet. Im folgenden wird die Untersuchung aus ökonomischen Gründen deshalb im wesentlichen auf Mittelwertspektren konzentriert.

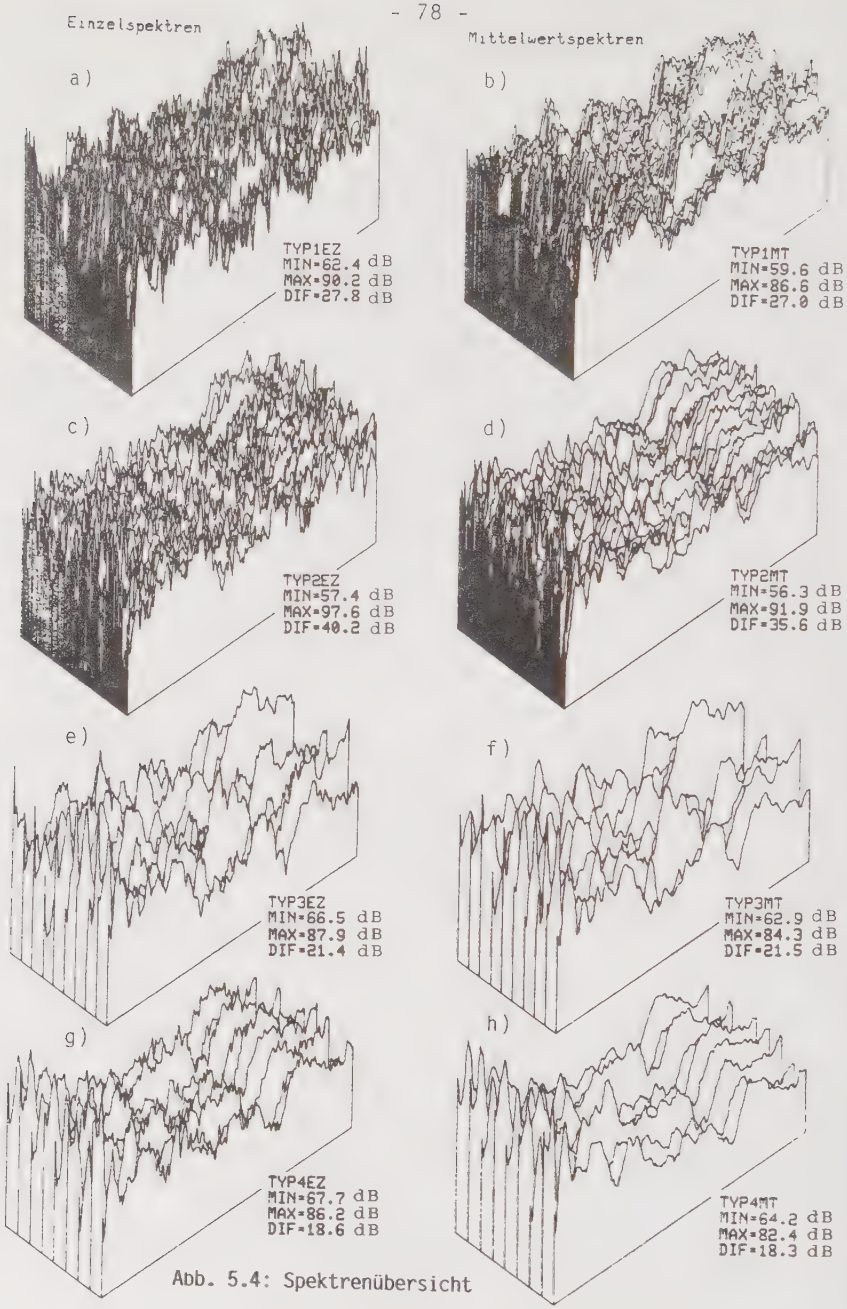


Abb. 5.4: Spektrenübersicht

5.2.2 Typ1: Mittelung über Kontexte

Zur Klärung der Frage, wie die Spektren von Sprecher zu Sprecher und von Realisation zu Realisation variieren, wurde zunächst über die Kontexte gemittelt. Aufgrund des vorhandenen Datenmaterials ergeben sich wie in Abbildung 5.4(a-b) dargestellt für 10 Sprecher je 10 Spektren. Von rechts nach links zeigen die Abbildungen 5.4(a-b) die zehn Realisationen für Sprecher 1, für Sprecher 2, usw. Wie bereits aus Bild 5.4(b) hervorgeht, lassen sich in bestimmten Frequenzbereichen Unterschiede (in erster Linie von Sprecher zu Sprecher) erkennen, die genauer beschrieben werden müssen. In Abbildung 5.5 sind dazu die genannten Spektren in gewohnter Form übereinander dargestellt; d.h., jedes Bild zeigt für einen Sprecher 10 jeweils durch Mittelung über die neun Kontexte gewonnene Spektren. Wie leicht zu erkennen ist, weichen die Spektren innerhalb der Sprecher nur geringfügig voneinander ab (etwa 3dB, siehe Abschnitt 5.4): dagegen ergeben sich deutliche Unterschiede von Sprecher zu Sprecher. So ist das bei den meisten Sprechern vorhandene Maximum bei 1.7 kHz bei den Sprechern 3 und 5 nur schwach ausgeprägt. Insgesamt auffallend ist die große Konsistenz, mit der einzelne Sprecher ihre /f/-Laute (immerhin in sehr heterogenen lautlichen Umgebungen) realisieren. Bezogen auf das hier angestrebte Meßverfahren zur nachträglichen Bestimmung der Eigenschaften von Übertragungskanälen ist festzustellen, daß /f/-Laute nicht von allen Sprechern in derselben Weise realisiert werden, im Spektrum jedoch soviele Übereinstimmungen aufweisen, daß ihre Anwendbarkeit als Eichsignal durchaus infrage kommt, sofern die Spektren über eine genügende Anzahl von Sprechern gemittelt werden und die dann gegebene Unsicherheit quantitativ bekannt ist.

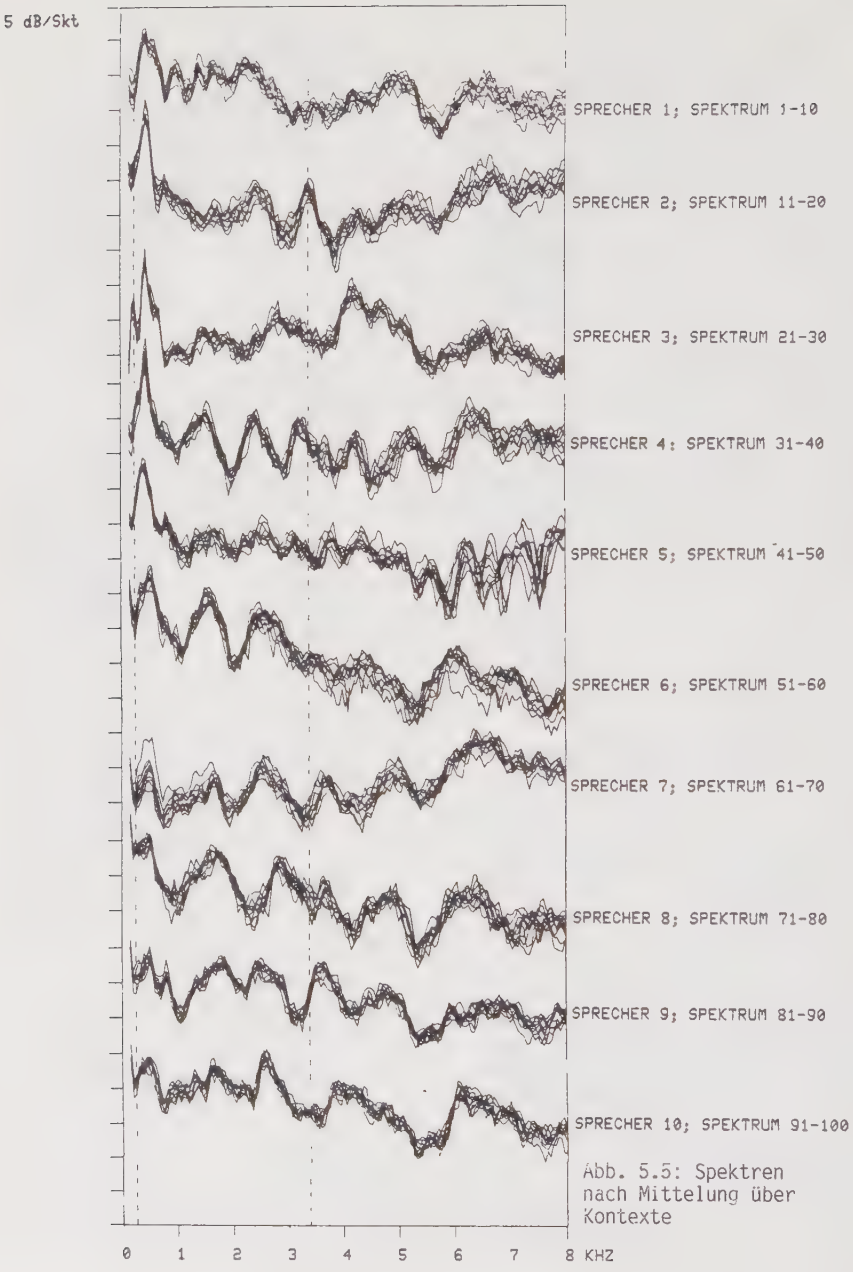


Abb. 5.5: Spektren nach Mittelung über Kontexte

5.2.3 Typ2: Mittelung über alle Realisationen

Das in 5.2.2 gewonnene Ergebnis wirft nun die Frage auf, ob durch Beschränkung auf einen einzelnen Kontext die Sprechervariabilität in den Spektren reduziert werden kann. Es wurde daher für jeden Sprecher und für jeden Kontext getrennt über alle Realisationen gemittelt, so daß sich insgesamt 90 Spektren ergeben. Diese sind in Abbildung 5.4(c-d) derart sortiert, daß von rechts nach links die Spektren für den 1. Kontext (/Vater/) von Sprecher 1 bis 10, dann für den 2. Kontext (/schläft viel/) von Sprecher 1 bis 10, usw. dargestellt sind. (Zur Beschreibung der Kontexte siehe Abschnitt 4.2.5) Unmittelbar voneinander abweichende Strukturen sind aus der Abbildung allein nicht zu erkennen. Daher wurden in Abbildung 5.6 analog zu Abbildung 5.5 für die Kontexte getrennt die Spektren für die einzelnen Sprecher, gemittelt über alle Realisationen dargestellt. Dabei zeigt sich, daß verschiedene Sprecher die /f/-Laute in den unterschiedlichen Kontexten auch deutlich unterschiedlich realisieren (zu den Streuungen im einzelnen siehe Abschnitt 5.4).

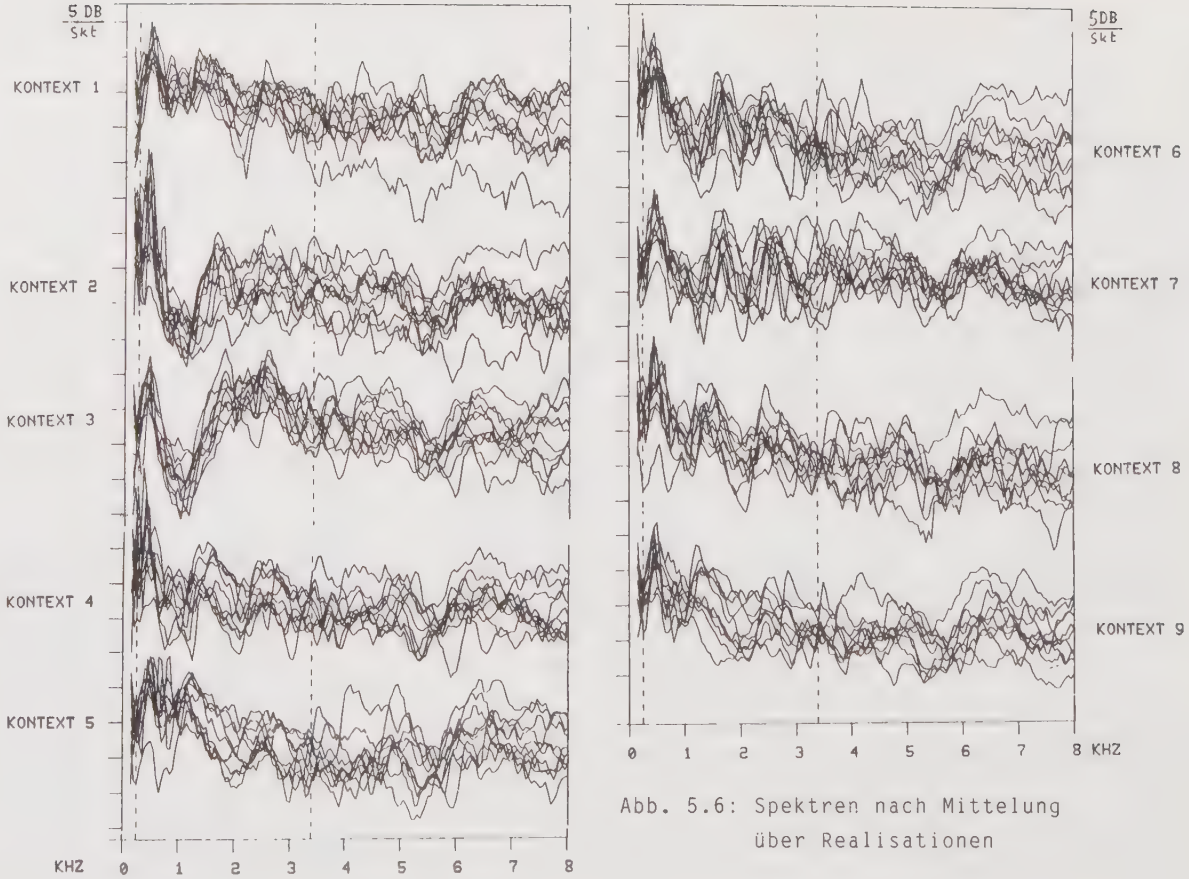


Abb. 5.6: Spektren nach Mittelung über Realisationen

5.2.4 Typ3: Mittelung über alle Realisationen und Kontexte

Mithilfe von Mittelungstyp 3 sollen die rein sprecherspezifischen Unterschiede festgestellt werden. Es wurde daher für jeden Sprecher über alle Realisationen und Kontexte gemittelt, so daß sich für jeden der 10 Sprecher ein Spektrum ergibt. Diese sind in Bild 5.4(e), rechts beginnend mit Sprecher 1 dargestellt. Wiederum sind zur besseren Veranschaulichung alle Spektren zusätzlich übereinander in den Abbildungen 5.7

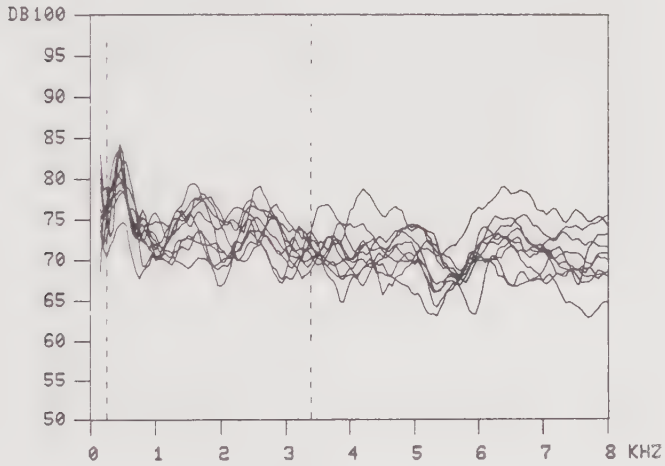


Abb. 5.7: Spektren nach Mittelung über alle Realisationen und Kontexte für jeden einzelnen Sprecher, überlagert dargestellt

und 5.8 wiedergegeben. Die Unterschiede zwischen den einzelnen Sprechern sind teilweise beträchtlich; so liegt z.B. bei Sprecher 2 bei 3.4 kHz ein Maximum, bei Sprecher 7 dagegen an annähernd derselben Stelle ein Minimum. Andererseits finden sich wieder viele Übereinstimmungen, wie das gemeinsame Maximum bei 1.7 kHz, das sicherlich auf die akustische Struktur des /f/ zurückzuführen ist.

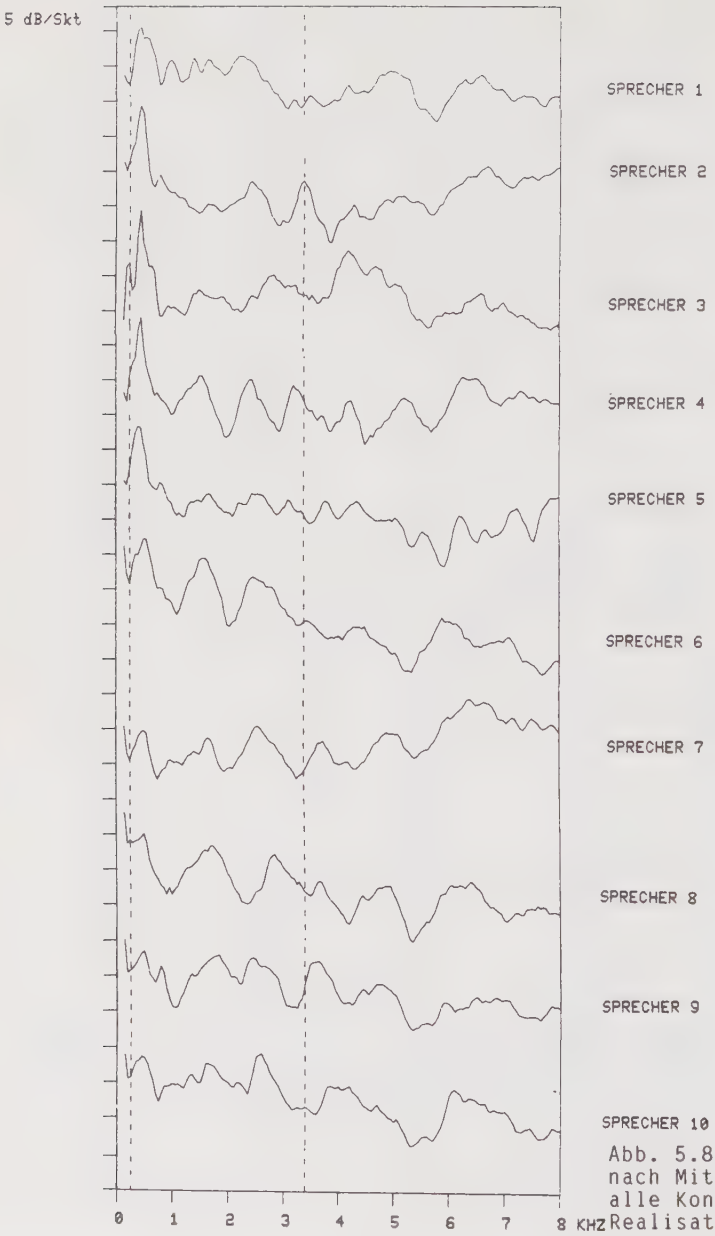


Abb. 5.8: Spektren nach Mittelung über alle Kontexte und Realisationen

5.2.5 Typ4: Mittelung über alle Sprecher und Realisationen

Bei Mittelung über alle Sprecher und Realisationen ergeben sich insgesamt neun Spektren, von denen jedes den /f/-Laut in einem der untersuchten Kontexte repräsentiert. Die Ergebnisse sind in den Abbildungen 5.4(g) für Einzelspektren, in 5.4(h) für

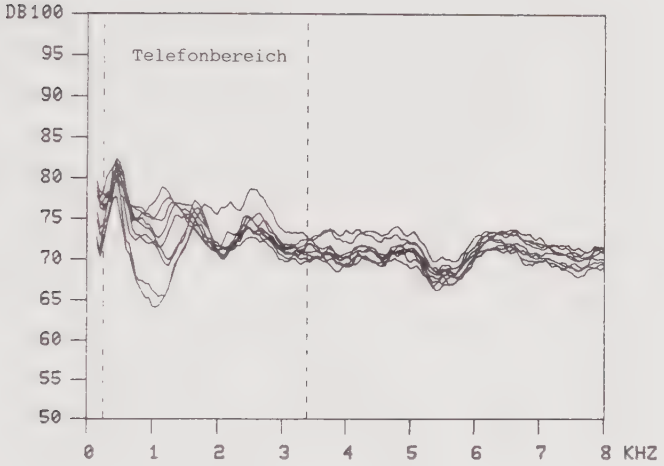


Abb.5.9: Spektren nach Mittelung über alle Sprecher und Realisationen für jeden Kontext getrennt, überlagert dargestellt.

Mittelwertspektren und in den Abbildungen 5.9 und 5.10 überlagert bzw. untereinander dargestellt. Es zeigt sich, daß gerade im Telefonbereich beträchtliche kontextbedingte Abweichungen im Spektrum vorkommen.

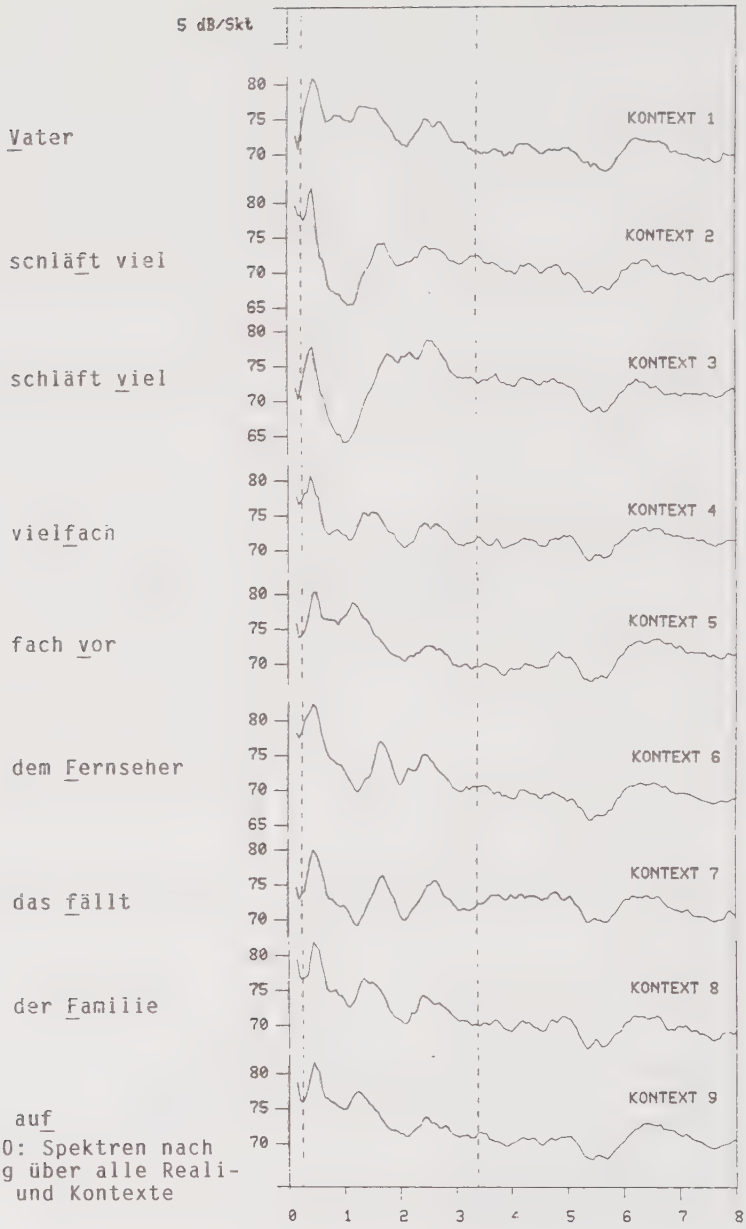


Abb. 5.10: Spektren nach Mittelung über alle Realisationen und Kontexte

5.2.6 Typ5: Mittelung über alle Sprecher, Realisationen und Kontexte

Die Abbildungen 5.11 und 5.12 zeigen das Ergebnis einer Mittelung über alle Sprecher, alle Kontexte und Realisationen für Einzel- und Mittelwertspektren, die wegen der sich hier ergebenden hohen Zahl an Mittelungen fast gleich sind. Als besondere Charakteristika der Spektren fallen im Telefonbereich die Maxima bei etwa 400, 1700 und 2400 Hz auf. Diese Formanten dürften von dem hinter der labio-dental gebildeten Enge gelegenen Teil des Vokaltraktes erzeugt werden. Die folgende Überlegung stützt diese Annahme: Im Viereck der Kardinalvokale entspricht ein Laut mit den oben genannten Formanten etwa dem /ə/. Es läßt sich leicht nachvollziehen, daß die Konfiguration des Vokaltraktes bei der Bildung des /f/ mit Ausnahme des labio-dentalen Artikulationsortes weitgehend derjenigen für das /ə/ entspricht.

Es sei an dieser Stelle darauf hingewiesen, daß für kriminaltechnische Anwendungen die hier diskutierten sowie die neun in Abschnitt 5.2.5 beschriebenen Spektren die größte Bedeutung besitzen. Im allgemeinen werden nämlich im Anwendungsfall nicht genau dieselben Kontextkonfigurationen vorliegen wie in dieser Untersuchung. Wenn dann /f/-Spektren verschiedenen Kontexten entnommen und gemittelt werden, so ist in der Regel das Spektrum des Typs 5 die Grundlage für die Ermittlung der Frequenzeigenschaften des Telefonkanals. Wenn dagegen alle Spektren derselben Kontextkategorie entstammen, ist das entsprechende Spektrum aus Typ4 (Mittelung über alle Sprecher und Realisationen) zu verwenden.

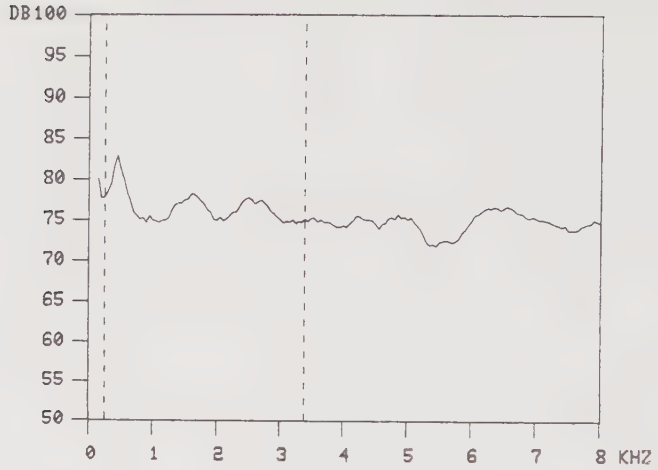


Abb. 5.11: Einzelspektrum, gemittelt über alle Sprecher, Kontexte und Realisationen

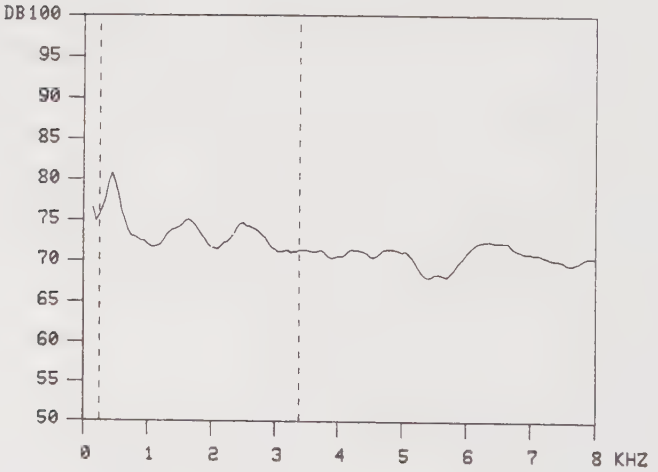


Abb. 5.12: Mittelwertspektrum, gemittelt über alle Sprecher, Kontexte und Realisationen

5.3 Einfluß des lautlichen Kontextes

5.3.1 Segmentation der Lautumgebungen

Im vorhergehenden Abschnitt wurden die Spektralstrukturen von /f/-Realisierungen aufgezeigt, ohne näher auf die Problematik intrinsischer bzw. kontextuell bedingter Eigenschaften einzugehen. Dieser Frage soll in diesem Abschnitt nachgegangen werden. Dazu wurden in den Abbildungen 5.13 bis 5.20 zu jedem /f/-Spektrum (in Form des Mittelwertspektrums) die Spektren des vorhergehenden und nachfolgenden Lautes über bzw. unter dem /f/-Spektrum dargestellt. Bei den /f/-Spektren handelt es sich um die Mittelungen über alle Sprecher und Realisation (TYP4). Die Kontextspektren dagegen wurden nur von einem Sprecher (Sprecher 1) berechnet. Die Segmentationsgrenzen für die Umgebungen erfassen nicht den gesamten vorhergehenden und nachfolgenden Laut sondern wurden vom Beginn bzw. Ende des /f/ bis zum Beginn bzw. Ende der stationären Phase des Kontextlautes gewählt. Auf diese Weise sollte vermieden werden, daß die Messung vermeidbar durch Spektralstrukturen der übernächsten Laute beeinflußt werden.

Die Segmentation wurde - wie bereits erwähnt - an einem graphischen Bildschirm durchgeführt mit der Möglichkeit, die segmentierten Abschnitte unmittelbar anzuhören und die Segmentation anschließend zu verfeinern. Die Segmentgrenzen wurden automatisch in Dateien abgelegt, auf die die Programme zur Berechnung der Spektren automatisch zugreifen konnten.*)

---

*)Für die Segmentation wurde bis zur Fertigstellung eines eigenen komfortablen Segmentationssystems, das an einem RAMTEK Farb-Raster-Bildverarbeitungssystem abeitet, ein beim Bundeskriminalamt vorhandenes Programmpaket (für TEKTRONIX Speicherbildschirm) benutzt.

5.3.2 Kontext 1: /f/ in /Vater/

Das /f/ in /Vater/ in initialer Position war in allen Fällen wegen des stark unterschiedlichen Signalcharakters von [f] und [a] leicht zu segmentieren. Vom nachfolgenden [a] ist regressiv ein koartikulatorischer Einfluß zu erwarten. Abbildung 5.13 zeigt ein relativ ausgeglichenes /f/-Spektrum, leichte Maxima sind an den Formantpositionen des [a] zu erkennen. Dabei kann es sich durchaus um Auswirkungen der Koartikulation, aber auch um eine /f/-typische Formantstruktur handeln. Die Tatsache, daß die Lage der Maxima nicht immer exakt übereinstimmen, sollte hier nicht zu stark bewertet werden, denn wie bereits geschildert stammt das /a/-Spektrum von nur einem Sprecher, während das /f/-Spektrum aus vielen Realisationen mehrerer Sprechern gemittelt wurde. Dieser Umstand ist auch bei den Erörterungen der folgenden Abschnitte zu bedenken.

Ein Vergleich mit FANTs Messungen des /fa/ (Abschnitt 3.2.1, Abbildung 3.1) bietet sich an. Dabei ist zu berücksichtigen, daß das /f/-Spektrum in Abbildung 5.13 aus vielen Einzelmessungen gemittelt wurde und allein schon deshalb einen flacheren Verlauf aufweisen muß als das nur einer Realisierung entstammende Spektrum FANTs. Eine Zuordnung der Extrema ist wegen des weitgehend ausgeglichenen Kurvenverlaufs in der eigenen Messung kaum möglich. Wird FANTs Kurve II zum Vergleich herangezogen, fällt allerdings das übereinstimmende Maximum bei etwa 2.5 kHz auf. Abbildung 5.13 läßt vermuten, daß die Ursache in einem koartikulatorischen Einfluß des folgenden /a/ zu sehen ist. Dagegen spricht aber, daß dasselbe Maximum auch in der für die Struktur des /f/ relevantesten Messung, Abbildung 5.12 (Mittelung über alle Sprecher, Kontexte und Realisationen), zu beobachten ist und daher wohl als Formant des /f/ betrachtet werden darf.

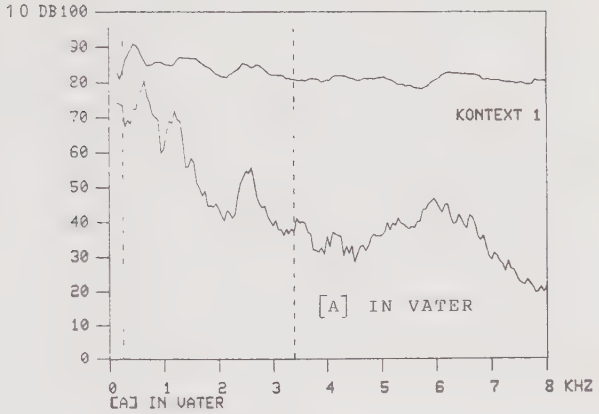


Abb. 5.13:
Spektrum des /f/ in /Vater/ und Kontextspektrum

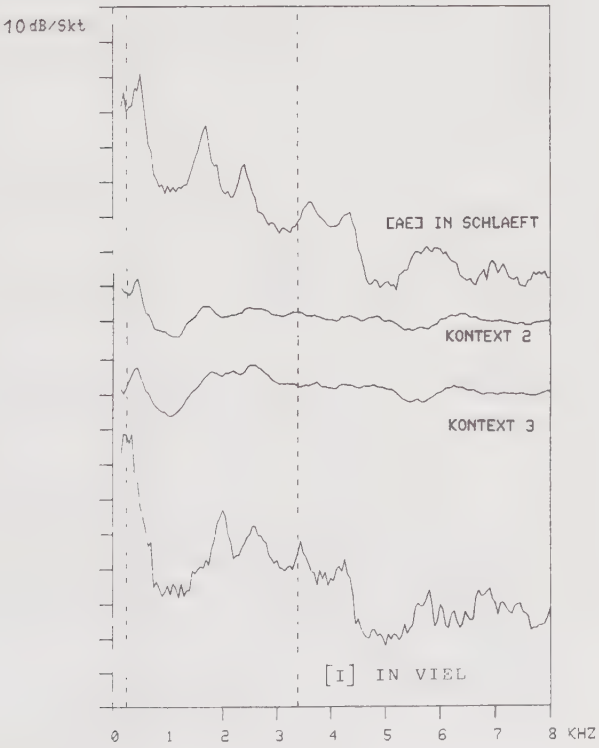


Abb. 5.14:
Spektrum des /f/ in /schläft viel/ und Kontextspektren

5.3.3 Kontexte 2 und 3: /f/ in /schläft viel/

Wie Abbildung 5.14 zeigt, weisen die Spektren beider /f/-Realisationen in /schläft viel/ einen weitgehend übereinstimmenden Verlauf auf, der in seiner Struktur offensichtlich starke Kontexteinflüsse vom vorhergehenden [ɛ:] bzw. nachfolgenden [i:] enthält. Die Segmentierbarkeit erschien nicht besonders schwierig, jedoch zeigte die auditive Kontrolle, daß auch bei einer visuell sauber erscheinenden Trennung der Signalteile von [f] und [i:] Ansätze des [i:] im [f] vorhanden waren. Das Energieminimum zwischen erstem und zweitem Formanten des [i:] und des [ɛ:] sind im /f/-Spektrum deutlich ausgeprägt wiederzufinden. Die ersten Formanten des [ɛ:] und [i:] verstärken das Maximum des /f/-Spektrums bei etwa 500 Hz, jedoch ist dieses Maximum bei allen /f/-Spektren zu beobachten und somit im wesentlichen wohl /f/-typisch und nicht auf koartikulatorischen Einfluß zurückzuführen.

5.3.4 Kontext 4: /f/ in /vielfach/

Ein besonders ausgeglichenes Spektrum zeigen die /f/-Realisationen in /fach/, Abbildung 5.15. Sein Verlauf entspricht im wesentlichen dem des durch Mittelung über alle 900 vorliegenden Realisationen gewonnenen Spektrums. Neben offensichtlich von vornherein geringen koartikulatorischen Einflüssen war hier auch die Segmentation besonders einfach durchzuführen. Die Übereinstimmung mit dem /f/-Spektrum in /Vater/, Abbildung 5.13, ist gut; für den Vergleich mit dem Spektrum FANTs für /fa/ gilt das in Abschnitt 5.3.2 Gesagte. Das /f/ in dieser lautlichen Umgebung ist daher aus koartikulatorischer Sicht besonders gut geeignet für eine spätere Anwendung im Sinne der vorliegenden Untersuchung.

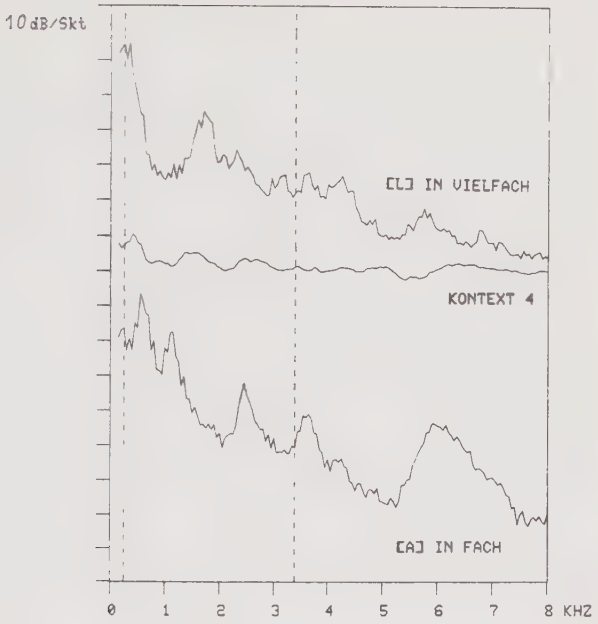


Abb. 5.15:
Spektrum des /f/ in /vielfach/ und Kontextspektren

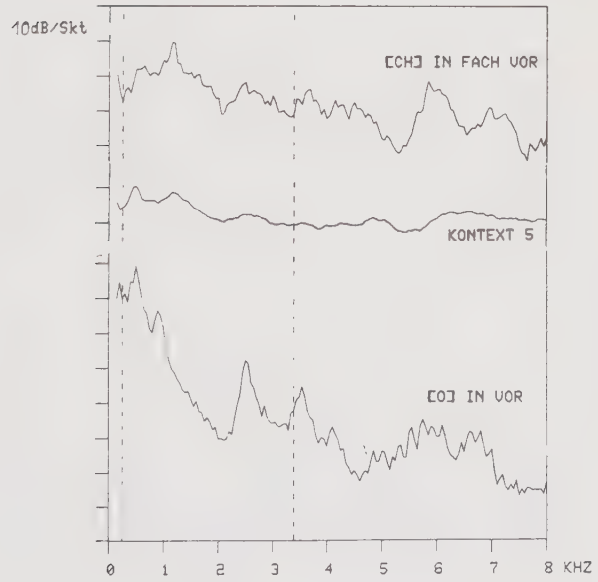


Abb. 5.16:
Spektrum des /f/ in /fach vor/ und Kontextspektren

5.3.5 Kontext 5: /f/ in /fach vor/

Für das /f/ in /fach vor/ ergibt sich ebenfalls ein relativ ausgeglichenes Spektrum, Abbildung 5.16. Dies war nicht unbedingt zu erwarten, da einerseits wegen des ähnlichen Signalcharakters des [f] und des vorhergehenden [x] im Oszillogramm teilweise Probleme bei der Segmentation auftraten und andererseits das [x] wegen seines weiter hinten liegenden Artikulationsortes eine deutlich ausgeprägte Formantstruktur besitzen kann. Auf dem Bildschirm war die Trennung von [x] und [f] schwierig, bei Hinzunahme der auditiven Kontrolle der Lautsegmente ließ sich jedoch meist eine recht gute Segmentation erreichen. Hinzu kommt, daß, wie Abbildung 5.16 bestätigt, das [x] hier keine stark ausgeprägten - und vermutlich auch bezüglich verschiedener Sprecher und Realisationen nicht stabile - Formanten besitzt. Eine regressive Koartikulation ist ebenfalls nicht zu erkennen.

5.3.6 Kontext 6: /f/ in /dem Fernseher/

Auch für das /f/ in /dem Fernseher/ ist kein deutlicher Kontexteinfluß zu erkennen, Abbildung 5.17. Obgleich sowohl das vorhergehende [m] wie auch das nachfolgende [e] in der vorliegenden Realisierung einen Formanten bei etwa 1.4 kHz besitzen, liegt an dieser Stelle in den gemittelten /f/-Spektren ein Minimum.

5.3.7 Kontext 7: /f/ in /das fällt/

In dieser lautlichen Umgebung erweisen sich die in fast allen /f/-Spektren zu findenden Extrema bei 1.3 kHz (Minimum) und 1.8 kHz (Maximum) als besonders stark ausgeprägt, Abbildung 5.18. Zumindest im vorliegenden Beispiel kann dies damit erklärt

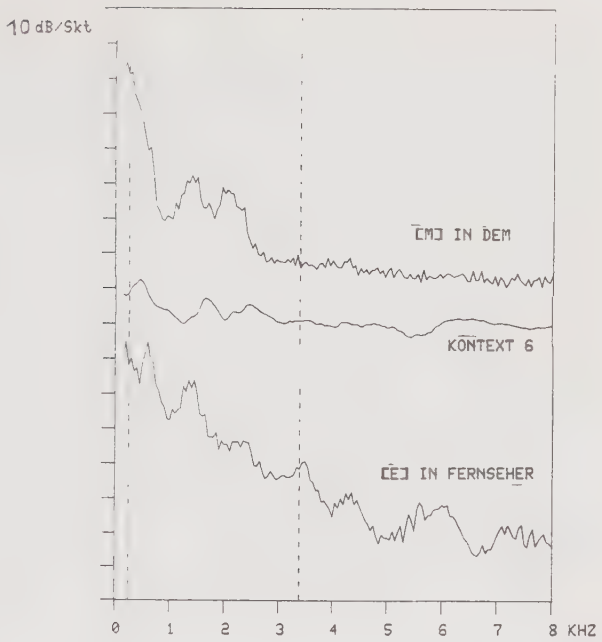


Abb. 5.17:
Spektrum des /f/ in /Fernseher/ und Kontextspektren

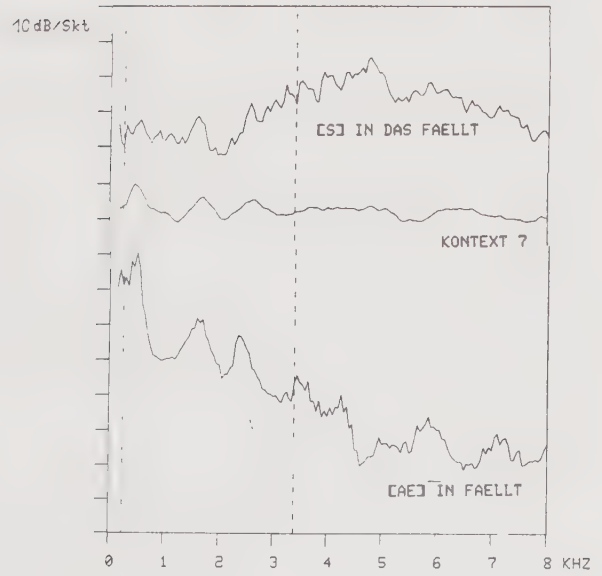


Abb. 5.18:
Spektrum des /f/ in /das fällt/ und Kontextspektren

werden, daß sowohl [s] wie auch [ɛ] einen Formanten bei 1.8 kHz und das [ɛ] zusätzlich ein Minimum bei 1.3 kHz aufweisen. Die Segmentation zwischen [s] und [f] war relativ schwierig und konnte nur mit ausgeprägter auditiver Kontrolle hinreichend gut durchgeführt werden.

5.3.8 Kontext 8: /f/ in /der Familie/

Das /f/-Spektrum aus dem Kontext /der Familie/, Abbildung 5.19, zeigt wieder einen relativ ausgeglichenen Verlauf, der sehr gut mit denen aus den anderen Messungen der /f/-Spektren mit nachfolgendem /a/, Abbildungen 5.13 und 5.15,übereinstimmt. Wegen der unterschiedlichen Artikulationsweisen des /r/ könnten in anderen Fällen allerdings andere Verhältnisse vorliegen.

5.3.9 Kontext 9: /f/ in /auf/

Kontext 9 zeigt das /f/ in finaler Position in /auf/, Abbildun 5.20. Die Segmentation war in allen Fällen sehr einfach; das /f/-Spektrum zeigt die übliche Struktur mit vergleichsweise schwach ausgeprägten Minima und Maxima. Die /f/-Laute in initialer und finaler Position haben für die vorliegende Untersuchung insofern besondere Bedeutung, als sie kontextuell nur einseitig beeinflußt werden und in finaler Position außerdem die größten Dauern aufweisen. Letzteres ergibt für die Mittelwertspektren eine besonders große Zahl zu mittelnder Spektren und damit einen besonders glatten Kurvenverlauf.

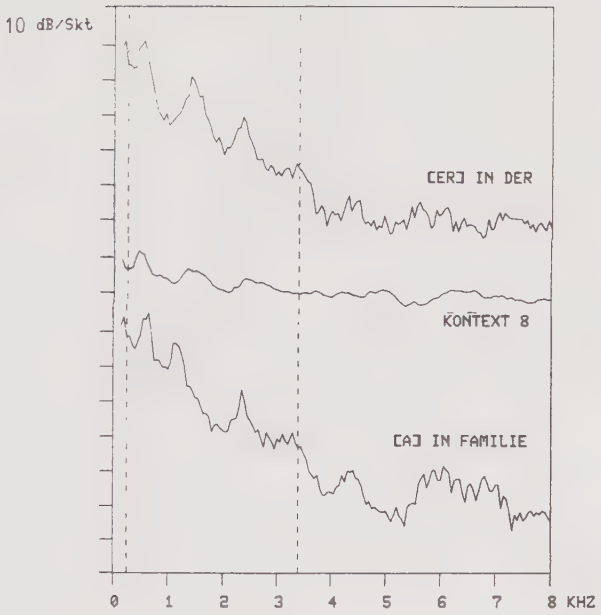


Abb. 5.19:
Spektrum des /f/ in
/Familie/ und
Kontextspektren

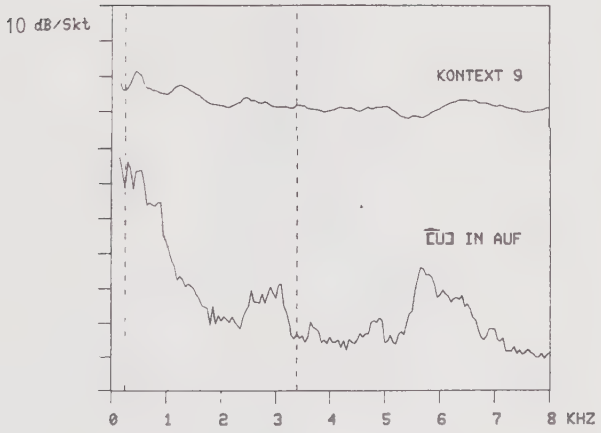


Abb. 5.20:
Spektrum des /f/ in
/auf/ und
Kontextspektrum

5.4 Streubreiten der /f/-Spektren

5.4.1 Berechnung der Streubreiten

Wie in Abschnitt 5.2 gezeigt wurde, weisen die /f/-Spektren unter den verschiedenen Randbedingungen unterschiedliche Strukturen auf. Für die spätere Anwendung ist nun wichtig zu wissen, mit welchen Streubreiten (gleichzusetzen mit der Unsicherheit bei der Anwendung) zu rechnen ist, wenn aus dem zur Verfügung stehenden Fallmaterial 1 bis N /f/-Laute aus 1 bis N Kontexten segmentiert werden können. Für jede dieser Konfigurationen muß für jeden Frequenzwert ein Intervall angebbar sein, das ein Maß für die Unsicherheit bei der Bestimmung der Übertragungsfunktion darstellt. Statistische Maße für eine solche Angabe, z.B. Toleranzintervalle, die angeben, mit welcher Wahrscheinlichkeit wieviel Prozent der Meßdaten in einem vorgegebenen Intervall liegen, sind wegen der Voraussetzung einer Normalverteilung nicht ohne weiteres anwendbar. Es wurde daher von den für jeden Frequenzwert berechneten Amplitudenhistogrammen versuchsweise beidseitig eine vorgegebene Prozentzahl der Meßdaten (mit einer durch die Histogrammberechnung vorgegebenen Ungenauigkeit) abgeschnitten. Diese Vorgehensweise wurde wieder verworfen, weil sie auf der Basis der vorhandenen Datenmenge trotz der großen Zahl der Spektren eine zu große Ungenauigkeit aufwies. Als Maß für die Streubreiten wurde schließlich die Differenz zwischen größter und kleinster Amplitude sämtlicher Spektren an jedem Frequenzwert festgelegt. Um selten auftretende oder durch Artefakte bedingte Ausreißer zu beseitigen, wurden zusätzlich um die Extrema bereinigte Intervalle angegeben.

Wie bereits mehrfach geschildert, ist zu erwarten, daß die Spektren vom Typ Einzelspektrum größere Streuungen aufweisen als die Mittelwertspektren. Dies wird anhand der Abbildungen 5.21 und 5.22 deutlich. Die Abbildungen zeigen als Histogramm die Verteilung der gemessenen Amplituden für alle Sprecher, Realisationen und Kontexte eines jeden Spektrumtyps, also jeweils von 900

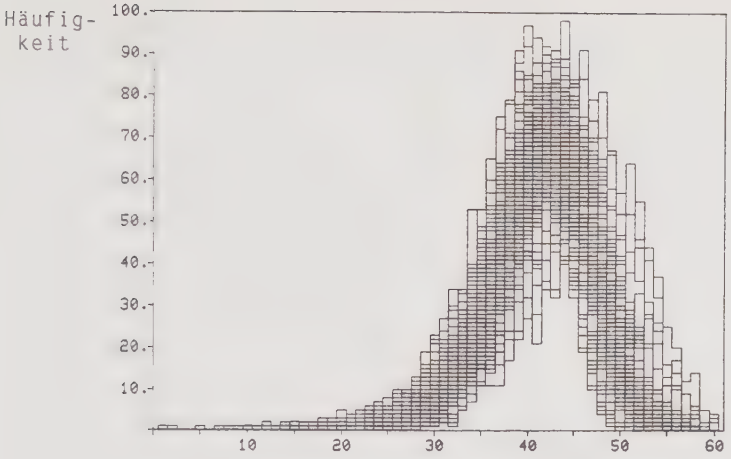


Abb. 5.21: Häufigkeitsverteilung der Amplituden der Einzelspektren (verteilt auf 60 äquidistante Kanäle)

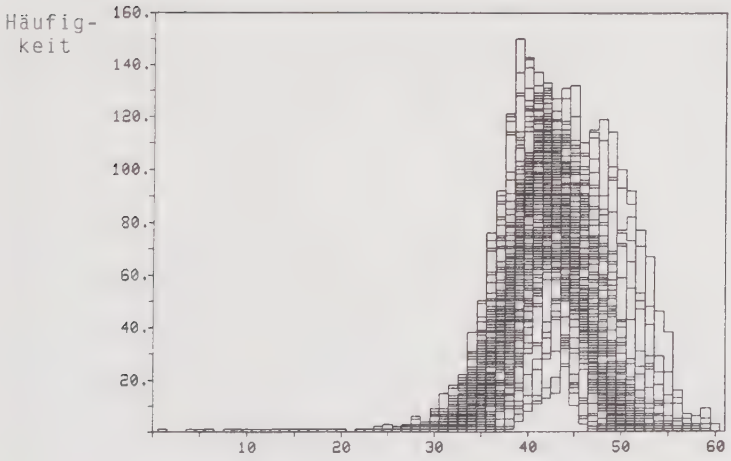
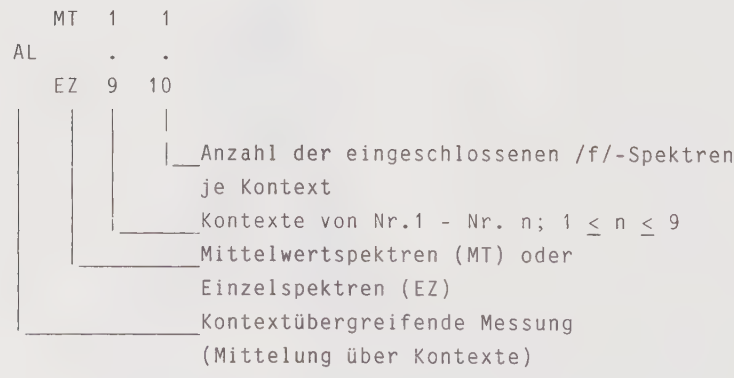


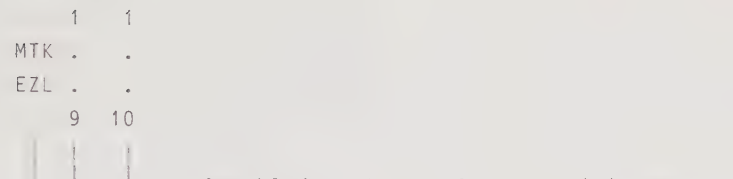
Abb. 5.22: Häufigkeitsverteilung der Amplituden der Mittelwertspektren (verteilt auf 60 äquidistante Kanäle)

Spektren. Dabei ist deutlich zu sehen, daß die Amplitudenverteilung der Mittelwertspektren einen schmaleren Verlauf aufweist.

Für die praktische Anwendung ergibt sich aufgrund der Notwendigkeit, den verschiedenen Bedingungen im Einzelfall (Zahl der /f/-Laute, Art der Kontexte, usw.) gerecht zu werden, eine große Menge von Daten, die einer überschaubaren Strukturierung bedarf. Aus der Art der Aufbereitung der vorhandenen Daten ergeben sich zwei "schlauch"-artige Begrenzungslinien um die durch /f/-Spektren gemessenen Übertragungsfunktionen, die als Maß für die Sicherheit bei der Messung dienen können. Dabei muß allerdings bedacht werden, daß diese Begrenzungen nicht den gesamten möglichen Fehler erfassen, sondern nur die durch unterschiedliche Spektren des /f/-Lautes bedingten Unsicherheiten berücksichtigt.

Tabelle 5.1 zeigt die Systematik der für jede Konfiguration bei der Messung abrufbaren insgesamt 360 Dateien, von denen jede zwei Intervalle enthält; diejenigen über alle Meßwerte und die extremwertbereinigten. Die Namensgebung wurde wie folgt gewählt:





Je nach Art und Menge des zur Verfügung stehenden Sprachmaterials kann also bei der Messung auf 720 Datensätze mit Streubreiten zurückgegriffen werden. Auf ALMT11.GRZ und ALMT19.GRZ ist beispielsweise zurückzugreifen, wenn aus Kontext 1 /f/-Spektren einmal bzw. zehn mal vorliegen. ALMT91.GRZ und ALMT910.GRZ sind anzuwenden, wenn Spektren aus den Kontexten 1-9 je einmal bzw. je zehn mal vorliegen. Diese Auswahl der Mittelung von Untermengen der Spektren aus verschiedenen Kontexten ist im Grunde willkürlich gibt aber einen gewissen Aufschluß über die Verhältnisse nach der Zusammenfassung einzelner Spektren durch Mittelung.

Die Abbildungen 5.23 bis 5.26 zeigen die Begrenzungslinien der genannten Dateien. Abb. 5.23 ist zugleich ein Beispiel dafür, daß die vorgenommene Extremwertbereinigung sinnvoll und notwendig ist. Das untere Spektrum ist sicherlich ein "Ausreißer", der wahrscheinlich auf eine einzelne unkorrekte Segmentation zurückzuführen ist. Dadurch sind im Signalsegment erhebliche Anteile des vorhergehenden oder folgenden Lautes enthalten, die das Spektrum verfälschen.

Tabelle 5.1: Dateien der Streubreiten (Begrenzungslinien)

Dateiname	Kontext	Anzahl	Typ
ALMT11.GRZ	1	1	Mittel
ALMT12.GRZ	1	2	Mittel
.	.	.	.
ALMT110.GRZ	1	10	Mittel
ALMT21.GRZ	1-2	1	Mittel
ALMT22.GRZ	1-2	2	Mittel
.	.	.	.
ALMT910.GRZ	1-9	10	Mittel
MTK11.GRZ	1	1	Mittel
MTK12.GRZ	1	2	Mittel
.	.	.	Mittel
MTK110.GRZ	1	10	Mittel
MTK21.GRZ	2	1	Mittel
.	.	.	Mittel
MTK99.GRZ	9	9	Mittel
MTK910.GRZ	9	10	Mittel
ALEZ11.GRZ	1	1	Einzel
ALEZ12.GRZ	1	2	EInzel
.	.	.	.
ALEZ110.GRZ	1	10	Einzel
ALEZ21.GRZ	1-2	1	Einzel
.	.	.	.
ALEZ910.GRZ	1-9	10	Einzel
EZL11.GRZ	1	1	Einzel
EZL12.GRZ	1	2	Einzel
.	.	.	Einzel
EZL99.GRZ	9	9	Einzel
EZL910.GRZ	9	10	Einzel

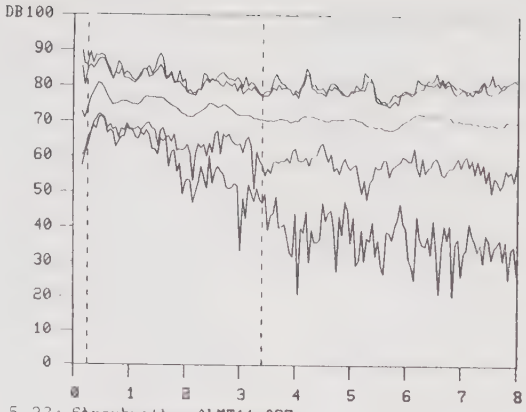


Abb. 5.23: Streubreiten ALMT11.GRZ

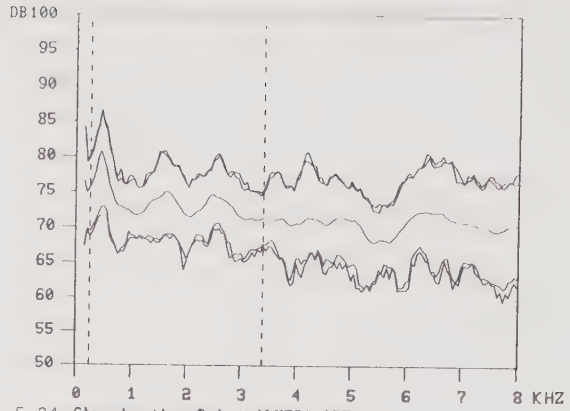


Abb. 5.24: Streubreiten Datei ALMT91.GRZ

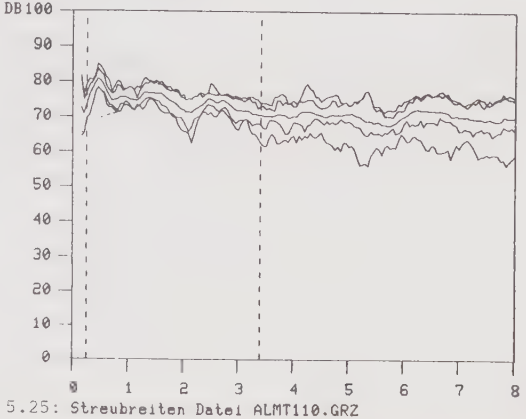


Abb. 5.25: Streubreiten Datei ALMT110.GRZ

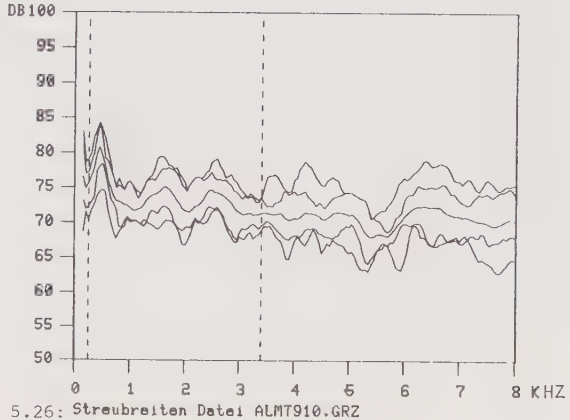


Abb. 5.26: Streubreiten Datei ALMT910.GRZ

5.4.2 Mittlere Streubreiten für unterschiedliche Mengen von /f/ Realisationen

Je mehr /f/-Realisierungen zur Verfügung stehen, desto geringer ist die Unsicherheit bei der späteren Messung der Übertragungsstrecke. Um die Frage nach einer notwendigen Mindestzahl von /f/-Realisierungen bzw. nach einer Sättigung der Konvergenz der Streubreiten mit zunehmender Anzahl beantworten zu können, wurden für jede Streubreitendatei der Mittelwert zwischen oberer und unterer Begrenzungslinie über alle Frequenzwerte bzw. nur für den Telefonkanal berechnet. Die Tabellen 5.2 bis 5.9 zeigen die Ergebnisse für Einzelspektren, die Tabellen 5.10 bis 5.17 in entsprechender Weise für Einzelspektren. In Tabelle 5.2 sind für den Spektrentyp Mittelwertspektrum für jeden Kontext getrennt die gemittelten Streubreiten in dB angegeben, wenn (von links nach rechts) ein, zwei, ..., zehn /f/-Laute dieses Kontextes im Sprachmaterial vorliegen. Die Tabellen 5.3 bis 5.5 enthalten die entsprechenden Ergebnisse, wenn nur der Frequenzbereich des Telefonkanals berücksichtigt und/oder wenn eine Extremwertbereinigung vorgenommen wird. Die Tabellen 5.6 bis 5.9 (und entsprechend Tabellen 5.14 bis 5.17 bei den Mittelwertspektren) gelten für den Fall, daß die /f/-Realisierungen aus unterschiedlichen Kontexten stammen. Zeile 1 in Tabelle 5.6 zeigt die Werte, wenn ein /f/-Laut im Kontext 1 ein bis zehn Mal vorliegt, ist also identisch mit Zeile 1 in Tabelle 5.2. Zeile 2 zeigt die mittleren Streubreiten, wenn Kontext 1 und 2 je ein bis zehn Mal vorliegen, usw.

Ein Vergleich der Tabellen läßt unmittelbar erkennen, daß Mittelwertspektren im Vergleich mit Einzelspektren die erwarteten geringeren Streubreiten aufweisen. Dieser Unterschied wird geringer, wenn bei den Einzelspektren mehr Realisationen in den Mittelungsprozeß einbezogen werden. Es ist sogar wahrscheinlich - anhand der hier zur Verfügung stehenden Datenmengen allerdings nicht nachzuweisen - daß bei einer sehr hohen, für die spätere Anwendung aber unrealistischen) Zahl von /f/-Realisierungen Einzelspektren besser geeignet sind, da sich eine geringere

Kontextanfälligkeit dann auswirken müßte. Die geringsten Streubreiten finden sich in den Tabellen 5.16 und 5.17, also dann, wenn möglichst viele /f/-Laute aus möglichst vielen lautlichen Umgebungen zugrunde liegen.

Tabelle 5.12 ist zu entnehmen, daß keine wesentlichen Unterschiede in den Streubreiten bestehen, wenn /f/-Realisierungen aus einzelnen Kontexten verwendet werden, sofern nur der betreffende /f/-Laut genügend häufig im Untersuchungsmaterial auftritt. Für neun oder zehn /f/-Realisierungen einer bestimmten Kontextkategorie ergeben sich unabhängig von ihrem Typ Streubreiten von etwa +/- 3.3 dB. Wenn die /f/-Spektren aus verschiedenen Kontexten genommen werden (was in der Anwendung die Regel sein wird), ergibt sich eine Konvergenz auf einen Sättigungswert von etwa +/- 2.2 dB.

Bei diesen Überlegungen muß allerdings bedacht werden, daß die einzelnen Werte der Tabellen unter statistischem Aspekt unterschiedlich sicher sind. Aus einem Untersuchungsmaterial von 900 Spektren läßt sich der Fall, daß ein Kontext einmal vorliegt für jede Kontextkategorie 100 mal konstruieren (Spalte 1 in den kontextabhängigen Tabellen), daß er 10 mal vorliegt aber nur 10 mal (Spalte 10). Diese Überlegungen gelten entsprechend für die Tabellen gemischter Kontexte. Alle Zusammenstellungen der genannten Art wurden auf dem Computer aus den beschriebenen 900 Spektren konstruiert.

Um die Ergebnisse anschaulicher zu gestalten, wurden einige Werte aus den Tabellen in den Abbildungen 5.27 bis 5.40 graphisch dargestellt. Der Übersichtlichkeit wegen erfolgte teilweise eine Einschränkung bei der Darstellung auf jeweils die erste (schlechtesten Werte), fünfte und neunte (besten Werte) Zeile einer Tabelle. Auffällig ist der in fast allen Bildern auftretende "Knick" beim fünfmaligen Vorliegen der betreffenden Laute. Grund dafür sind die bereits erwähnten Konstruktionsmöglichkeiten für die Kontextkombinationen aus den 900 Spektren des Basismaterials. Die folgende Aufstellung zeigt, wie oft Kontext oder Kontext-

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	35.14	23.91	18.52	16.21	14.49	12.71	12.28	11.87	10.98	10.44
KON 2	38.09	25.43	20.25	17.93	16.88	13.77	13.23	13.22	13.31	13.29
KON 3	35.28	24.03	19.35	16.36	15.57	12.81	12.43	12.13	11.83	11.76
KON 4	34.69	23.54	18.99	15.98	15.04	12.14	12.21	11.81	11.76	11.50
KON 5	34.96	24.08	19.58	17.33	16.06	13.50	13.05	12.89	12.52	12.34
KON 6	36.04	24.48	19.81	17.19	16.09	13.71	13.33	13.19	12.88	12.58
KON 7	35.16	23.31	18.39	15.52	14.74	11.95	11.40	11.22	11.06	10.93
KON 8	36.37	24.53	20.09	17.49	16.33	14.13	13.79	13.36	12.99	12.61
KON 9	35.04	23.44	18.32	15.84	15.08	12.46	12.03	11.50	11.20	11.19

GEMITTELTE STREUBREITEN FUER EINZELSPEKTREN,
EZL1.BRT

Tabelle 5.2

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	35.02	23.05	17.85	15.42	13.67	11.66	11.04	10.74	10.11	9.65
KON 2	35.58	23.93	19.14	16.63	15.67	12.78	12.17	11.95	11.97	12.09
KON 3	34.67	23.44	19.40	16.55	15.82	12.91	12.80	12.52	12.37	12.13
KON 4	34.76	23.81	19.44	16.40	15.65	12.75	12.69	12.29	12.13	11.80
KON 5	34.45	23.33	19.22	17.45	16.01	13.51	13.22	13.27	12.65	12.48
KON 6	34.76	23.56	19.69	16.57	15.49	13.24	13.21	12.87	12.71	12.22
KON 7	36.19	25.01	19.94	17.02	16.61	13.52	12.91	12.83	12.63	12.66
KON 8	34.87	24.13	19.68	17.25	16.31	14.22	13.63	13.27	12.88	12.53
KON 9	33.75	23.43	18.46	16.10	15.03	12.46	11.98	11.70	11.50	11.29

GEMITTELTE STREUBREITEN FUER EINZELSPEKTREN, TELEFONKANAL,
EZL2.BRT

Tabelle 5.3

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	28.95	19.01	14.62	12.38	11.25	8.40	8.15	7.82	7.56	7.38
KON 2	32.44	21.13	16.52	13.49	13.22	9.04	8.38	8.80	8.84	8.56
KON 3	30.30	20.02	15.49	12.77	12.15	9.26	8.91	8.50	8.17	8.06
KON 4	29.18	19.39	15.47	12.41	11.75	8.16	8.07	7.72	7.56	7.47
KON 5	29.28	19.96	15.81	13.02	12.41	8.89	8.75	8.63	8.29	7.98
KON 6	30.08	20.07	16.17	13.34	12.64	9.00	8.71	8.61	8.39	8.16
KON 7	28.71	18.92	14.67	11.97	11.20	8.13	7.75	7.66	7.40	7.39
KON 8	30.09	20.03	16.09	13.26	12.35	8.56	8.23	7.98	7.66	7.63
KON 9	28.98	19.52	14.74	12.11	11.34	8.13	7.91	7.67	7.44	7.32

GEMITTELTE STREUBREITEN FUER EINZELSPEKTREN, EXTREMWERTBEREINIGT,
EZL3.BRT

Tabelle 5.4

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	28.39	18.17	13.88	11.55	10.40	7.96	7.75	7.47	7.23	7.16
KON 2	30.33	19.66	15.46	12.90	12.67	9.23	8.59	8.60	8.59	8.51
KON 3	30.07	19.66	15.98	13.33	12.56	9.52	9.59	9.17	9.01	8.80
KON 4	29.04	19.64	15.96	12.97	12.25	8.86	8.82	8.51	8.34	8.19
KON 5	28.79	19.73	15.75	13.23	12.55	9.22	9.16	8.91	8.57	8.23
KON 6	29.40	19.34	15.76	12.86	12.26	9.12	9.02	8.82	8.53	8.25
KON 7	29.50	19.97	15.79	13.18	12.52	9.06	8.62	8.54	8.11	8.25
KON 8	29.55	20.12	16.04	13.23	12.38	8.39	8.10	8.04	7.79	7.73
KON 9	28.23	19.40	14.60	12.09	11.26	8.08	7.87	7.88	7.87	7.66

GEMITTELTE STREUBREITEN FUER EINZELSPEKTREN, EXTREMWERTBEREINIGT,
TELEFONKANAL; EZL4.BRT

Tabelle 5.5

Tabellen 5.2 bis 5.5: Gemittelte Streubreiten für Einzelspektren

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	35.14	23.91	18.52	16.21	14.49	12.71	12.28	11.87	10.98	10.44
KON 1-2	26.23	18.58	14.99	13.56	12.59	11.03	10.59	10.41	10.04	9.89
KON 1-3	22.41	16.41	13.67	12.43	11.71	10.34	10.07	9.98	9.74	9.65
KON 1-4	20.10	15.26	13.02	11.78	11.18	9.89	9.77	9.67	9.51	9.38
KON 1-5	18.67	14.57	12.56	11.53	11.02	9.87	9.72	9.65	9.45	9.34
KON 1-6	18.03	14.16	12.38	11.38	10.94	9.83	9.68	9.63	9.46	9.32
KON 1-7	17.15	13.47	11.91	10.94	10.62	9.57	9.38	9.33	9.18	9.09
KON 1-8	16.69	13.24	11.80	10.90	10.60	9.66	9.46	9.40	9.22	9.12
KON 1-9	16.69	13.24	11.80	10.90	10.60	9.66	9.46	9.40	9.22	9.12

GEMITTELTE STREUBREITEN FUER EINZELSPEKTREN, GEMISCHTE KONTEXTE,
ALEZ1.BRT

Tabelle 5.6

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	35.02	23.05	17.85	15.42	13.67	11.66	11.04	10.74	10.11	9.65
KON 1-2	25.58	17.67	14.53	12.92	11.82	10.57	9.95	9.70	9.37	9.27
KON 1-3	21.51	15.45	13.14	11.80	11.07	9.96	9.68	9.48	9.29	9.14
KON 1-4	19.42	14.69	12.70	11.61	10.76	9.74	9.64	9.48	9.26	9.07
KON 1-5	18.14	13.97	12.12	11.34	10.62	9.62	9.58	9.49	9.24	9.10
KON 1-6	17.56	13.34	11.75	10.93	10.38	9.43	9.36	9.27	9.11	8.93
KON 1-7	16.82	13.02	11.57	10.73	10.36	9.41	9.27	9.22	9.04	8.94
KON 1-8	16.34	12.81	11.41	10.56	10.30	9.47	9.28	9.23	9.02	8.91
KON 1-9	16.34	12.81	11.41	10.56	10.30	9.47	9.28	9.23	9.02	8.91

GEMITTELTE STREUBREITEN FUER EINZELSPEKTREN, GEMISCHTE KONTEXTE,
TELEFONKANAL; ALEZ2.BRT

Tabelle 5.7

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	28.95	19.01	14.62	12.38	11.25	8.40	8.15	7.82	7.56	7.38
KON 1-2	22.34	15.36	12.20	10.44	9.77	7.42	7.08	6.96	6.84	6.46
KON 1-3	19.38	13.68	11.19	9.62	9.30	6.69	6.47	6.30	6.13	5.91
KON 1-4	17.43	12.83	10.64	9.18	8.93	6.30	6.15	6.03	5.89	5.74
KON 1-5	16.43	12.35	10.39	9.13	8.81	6.13	6.02	5.90	5.79	5.66
KON 1-6	15.73	12.03	10.24	9.14	8.79	6.22	6.12	5.99	5.86	5.75
KON 1-7	14.91	11.55	9.88	8.88	8.53	6.08	5.92	5.84	5.71	5.63
KON 1-8	14.56	11.44	9.92	8.97	8.58	6.05	5.89	5.79	5.68	5.60
KON 1-9	14.56	11.44	9.92	8.97	8.58	6.05	5.89	5.79	5.68	5.60

GEMITTELTE STREUBREITEN FUER EINZELSPEKTREN, GEMISCHTE KONTEXTE,
EXTREMWERTBEREINIGT; ALEZ3.BRT

Tabelle 5.8

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	28.39	18.17	13.88	11.55	10.40	7.96	7.75	7.47	7.23	7.16
KON 1-2	21.91	14.78	11.60	10.16	9.41	7.45	7.24	7.08	6.95	6.95
KON 1-3	18.72	13.11	10.79	9.28	9.03	6.74	6.64	6.49	6.33	6.31
KON 1-4	16.87	12.54	10.56	9.11	8.88	6.72	6.60	6.45	6.31	6.32
KON 1-5	15.98	12.09	10.39	9.09	8.78	6.67	6.55	6.42	6.26	6.23
KON 1-6	15.17	11.59	10.06	8.99	8.52	6.48	6.41	6.28	6.11	6.08
KON 1-7	14.55	11.30	9.87	8.90	8.47	6.47	6.36	6.27	6.13	6.08
KON 1-8	14.24	11.13	9.88	8.90	8.44	6.40	6.34	6.20	6.06	6.04
KON 1-9	14.24	11.13	9.88	8.90	8.44	6.40	6.34	6.20	6.06	6.04

GEMITTELTE STREUBREITEN FUER EINZELSPEKTREN, GEMISCHTE KONTEXTE,
EXTREMWERTBEREINIGT, TELEFONKANAL; ALEZ4.BRT

Tabelle 5.9

Tabellen 5.6 bis 5.9: Gemittelte Streubreiten für Einzelspektren

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	35.62	21.13	17.48	11.98	13.91	10.47	10.30	10.02	11.75	11.17
KON 2	23.87	18.10	15.71	14.12	13.88	11.92	11.87	11.78	11.85	11.89
KON 3	19.22	15.38	12.91	11.97	11.34	10.19	10.06	9.97	9.68	9.55
KON 4	18.53	14.57	12.87	11.43	11.36	9.66	9.71	9.72	9.75	9.62
KON 5	19.21	15.44	13.56	12.55	12.00	10.70	10.45	10.44	10.39	10.30
KON 6	21.07	16.54	14.47	13.39	12.98	11.85	11.66	11.60	11.48	11.36
KON 7	19.14	14.53	12.46	11.44	11.16	9.90	9.74	9.79	9.68	9.64
KON 8	22.13	16.94	14.75	13.56	13.01	12.02	11.79	11.64	11.48	11.26
KON 9	17.27	14.13	12.13	11.31	11.38	9.97	9.88	9.75	9.86	9.97

GEMITTELTE STREUBREITEN FUER MITTELWERTSPEKTREN, MTK1.BRT

Tabelle 5.10

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	24.46	14.22	11.63	9.37	9.35	7.85	7.59	7.56	7.65	7.45
KON 2	22.78	16.57	14.71	13.01	12.83	10.72	10.54	10.58	10.90	10.80
KON 3	17.86	13.93	11.68	10.75	10.01	8.89	8.76	8.70	8.51	8.41
KON 4	17.77	13.96	12.63	10.99	10.65	9.56	9.40	9.40	9.46	9.26
KON 5	17.76	14.40	12.30	11.58	10.74	9.50	9.35	9.41	9.29	9.17
KON 6	19.73	15.04	12.94	11.71	11.33	10.11	9.99	9.84	9.91	9.69
KON 7	20.00	15.79	13.51	12.72	12.29	11.17	10.92	11.13	10.96	10.98
KON 8	20.23	15.25	13.00	11.87	11.49	10.55	10.33	10.29	10.06	9.86
KON 9	16.61	13.82	11.61	10.72	10.80	9.58	9.34	9.29	9.53	9.53

GEMITTELTE STREUBREITEN FUER MITTELWERTSPEKTREN, TELEFONKANAL, MTK2.BRT

Tabelle 5.11

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	19.53	14.56	12.20	9.66	9.92	6.59	6.45	6.31	6.33	6.24
KON 2	20.64	15.47	12.96	11.12	11.08	7.20	6.87	6.90	6.73	6.82
KON 3	16.77	13.01	10.92	9.67	9.34	7.15	7.07	6.94	6.62	6.49
KON 4	16.08	12.58	10.75	9.35	9.18	6.23	6.27	6.31	6.29	6.27
KON 5	16.93	13.35	11.43	9.84	9.74	7.17	7.13	7.10	6.95	6.85
KON 6	18.21	14.22	12.45	11.09	10.98	7.62	7.61	7.57	7.53	7.54
KON 7	16.52	12.45	10.61	9.26	9.20	6.42	6.41	6.37	6.26	6.21
KON 8	18.53	14.36	12.31	10.94	10.58	7.17	6.99	6.93	6.88	6.78
KON 9	15.20	12.16	10.43	9.30	9.42	6.45	6.47	6.40	6.40	6.33

GEMITTELTE STREUBREITEN FUER MITTELWERTSPEKTREN, EXTREMWERTBEREINIGT, MTK3.BRT

Tabelle 5.12

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	16.54	11.48	9.30	7.61	7.40	5.41	5.24	5.22	5.32	5.28
KON 2	19.59	14.38	12.01	10.19	10.29	7.69	7.37	7.20	7.09	7.29
KON 3	15.60	11.84	9.98	8.91	8.33	6.43	6.32	6.29	6.06	6.01
KON 4	15.72	12.19	10.50	9.27	9.02	6.55	6.60	6.63	6.62	6.56
KON 5	15.82	12.36	10.60	8.87	8.82	6.72	6.65	6.68	6.50	6.39
KON 6	16.89	12.91	11.11	9.62	9.57	7.09	7.05	6.96	6.92	6.80
KON 7	17.54	13.55	11.63	10.37	10.39	7.17	7.24	7.32	7.22	7.23
KON 8	16.90	13.08	10.99	9.94	9.34	6.83	6.67	6.76	6.59	6.41
KON 9	14.51	11.57	10.01	8.88	9.01	6.66	6.62	6.46	6.67	6.58

GEMITTELTE STREUBREITEN FUER MITTELWERTSPEKTREN, EXTREMWERTBEREINIGT, TELEFONKANAL; MTK4.BRT

Tabelle 5.13

Tabellen 5.10 bis 5.13: Gemittelte Streubreiten für Mittelwertspektren

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	35.62	21.13	17.48	11.98	13.91	10.47	10.30	10.02	11.75	11.17
KON 1-2	21.74	14.44	12.67	10.48	10.59	9.15	8.91	8.86	9.22	9.02
KON 1-3	17.32	12.70	11.21	9.93	9.77	8.83	8.65	8.63	8.77	8.65
KON 1-4	15.30	11.96	10.66	9.74	9.54	8.67	8.59	8.58	8.66	8.57
KON 1-5	14.25	11.56	10.44	9.65	9.44	8.71	8.60	8.57	8.66	8.56
KON 1-6	13.51	11.27	10.26	9.61	9.41	8.75	8.64	8.62	8.69	8.60
KON 1-7	12.87	10.87	9.98	9.37	9.20	8.57	8.46	8.47	8.52	8.46
KON 1-8	12.77	10.87	10.09	9.52	9.33	8.76	8.64	8.65	8.71	8.63
KON 1-9	12.32	10.65	9.87	9.35	9.18	8.64	8.53	8.52	8.59	8.52

GEMITTELTE STREUBREITEN FUER MITTELWERTSPEKTREN, GEMISCHTE KONTEXTE, ALMT1.BRT

Tabelle 5.14

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	24.46	14.22	11.63	9.37	9.35	7.85	7.59	7.56	7.65	7.45
KON 1-2	16.29	11.68	10.30	9.23	8.77	7.73	7.47	7.45	7.39	7.34
KON 1-3	13.49	10.39	9.42	8.58	8.10	7.30	7.16	7.20	7.09	7.05
KON 1-4	12.34	10.09	9.01	8.40	7.98	7.31	7.21	7.27	7.20	7.16
KON 1-5	11.67	9.64	8.83	8.26	7.88	7.27	7.19	7.22	7.16	7.09
KON 1-6	11.09	9.33	8.60	8.06	7.73	7.20	7.10	7.12	7.10	7.04
KON 1-7	10.89	9.26	8.60	8.07	7.77	7.28	7.18	7.24	7.21	7.17
KON 1-8	10.81	9.22	8.59	8.11	7.80	7.37	7.27	7.33	7.30	7.25
KON 1-9	10.56	9.13	8.47	8.00	7.75	7.35	7.22	7.28	7.26	7.22

GEMITTELTE STREUBREITEN FUER MITTELWERTSPEKTREN, GEMISCHTE KONTEXTE, TELEFONKANAL; ALMT2.BRT

Tabelle 5.15

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	19.53	14.56	12.20	9.66	9.92	6.59	6.45	6.31	6.33	6.24
KON 1-2	15.51	11.87	10.24	8.41	8.64	6.00	5.83	5.72	5.87	5.76
KON 1-3	13.83	10.88	9.40	8.11	8.25	5.68	5.57	5.41	5.42	5.33
KON 1-4	12.85	10.35	9.13	8.06	8.13	5.37	5.33	5.22	5.23	5.16
KON 1-5	12.25	10.17	8.99	8.05	8.14	5.29	5.30	5.25	5.25	5.21
KON 1-6	11.99	10.06	9.06	8.17	8.24	5.53	5.54	5.49	5.49	5.43
KON 1-7	11.47	9.75	8.81	8.00	8.07	5.47	5.44	5.39	5.39	5.34
KON 1-8	11.49	9.85	8.93	8.19	8.24	5.45	5.42	5.37	5.35	5.29
KON 1-9	11.18	9.68	8.80	8.11	8.18	5.38	5.37	5.31	5.27	5.23

GEMITTELTE STREUBREITEN FUER MITTELWERTSPEKTREN, GEMISCHTE KONTEXTE, EXTREMWERTBEREINIGT; ALMT3.BRT

Tabelle 5.16

	ANZAHL DER LAUTE									
	1	2	3	4	5	6	7	8	9	10
KON 1	16.54	11.48	9.30	7.61	7.40	5.41	5.24	5.22	5.32	5.28
KON 1-2	13.85	10.00	8.46	7.17	7.11	5.75	5.43	5.39	5.33	5.34
KON 1-3	11.83	9.10	7.71	6.78	6.67	5.27	5.08	4.96	4.78	4.82
KON 1-4	10.91	8.79	7.71	6.89	6.81	5.24	5.12	5.05	4.90	4.91
KON 1-5	10.36	8.63	7.58	6.84	6.73	5.14	5.05	5.02	4.94	4.94
KON 1-6	10.02	8.39	7.49	6.79	6.72	5.13	5.10	5.08	5.01	4.97
KON 1-7	9.87	8.33	7.49	6.85	6.83	5.27	5.21	5.20	5.16	5.12
KON 1-8	9.80	8.41	7.51	6.96	6.94	5.29	5.26	5.24	5.18	5.13
KON 1-9	9.61	8.33	7.46	6.96	6.95	5.24	5.22	5.20	5.15	5.12

GEMITTELTE STREUBREITEN FUER MITTELWERTSPEKTREN, GEMISCHTE KONTEXTE, EXTREMWERTBEREINIGT, TELEFONKANAL; ALMT4.BRT

Tabelle 5.17

Tabellen 5.14 bis 5.17: Gemittelte Streubreiten für Mittelwertspektren

kombinationen daraus im Sinne der Untersuchung zusammengestellt werden können. Ausgehend von 100 Realisationen des Testsatzes ergibt sich am Beispiel des Kontextes 1:

Kontext soll als Berechnungsgrundlage vorliegen von jedem Sprecher	Im Hintergrundmaterial ist diese Konfiguration enthalten	Aus methodischen Gründen unberücksichtigt bleiben vom Hintergrundmaterial
1 mal	100 mal	0 Spektren
2	50	0
3	30	10
4	20	20
5	20	0
6	10	40
7	10	30
8	10	20
9	10	10
10	10	0

Die Zusammensetzungen der einzelnen Kombinationen wurde so durchgeführt, daß möglichst viele Beiträge von möglichst vielen Sprechern stammen. Aus der Aufstellung wird unmittelbar klar, daß bei drei oder viermaligem Vorliegen des /f/ aus dem betrachteten Kontext zunehmend weniger, beim Übergang von vier nach fünf Realisierungen dagegen wieder alle Spektren in die Berechnung eingehen und die mittleren Streubreiten sich dann wieder vergrößern. Daraus ist zu folgern, daß die Zahlenwerte für diejenigen Kombinationen, die nicht das gesamte Untersuchungsmaterial umfassen, eher zu niedrig sind.

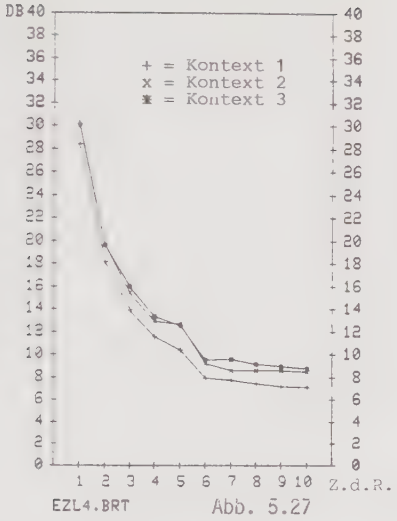


Abb. 5.27

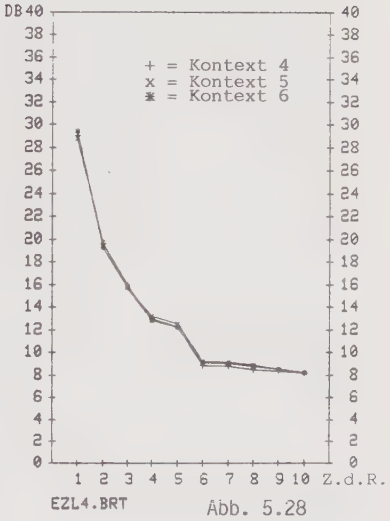


Abb. 5.28

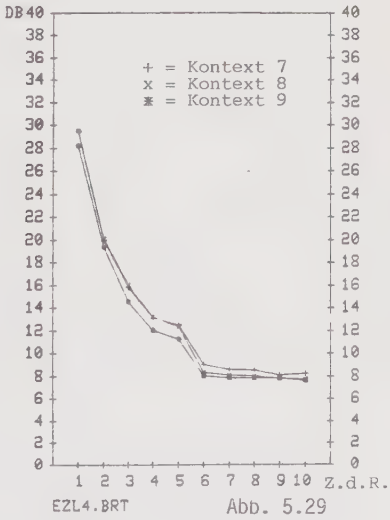


Abb. 5.29

Abb. 5.27 - 5.29:
Streubreiten für Einzelspektren nach Kontexten getrennt in Abhängigkeit von der Zahl der Realisierungen
(Daten aus Tabelle 5.5)

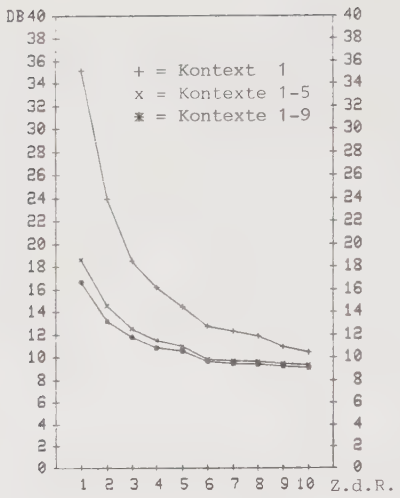


Abb. 5.30: ALEZ1.BRT

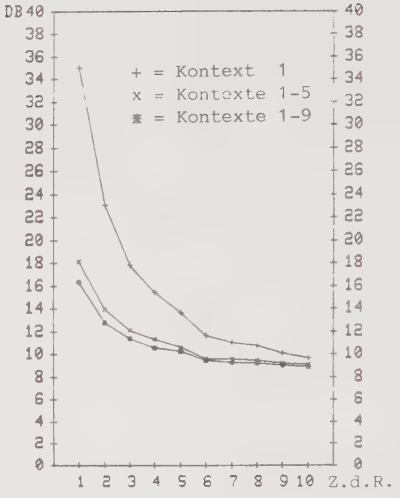


Abb. 5.31: ALEZ2.BRT

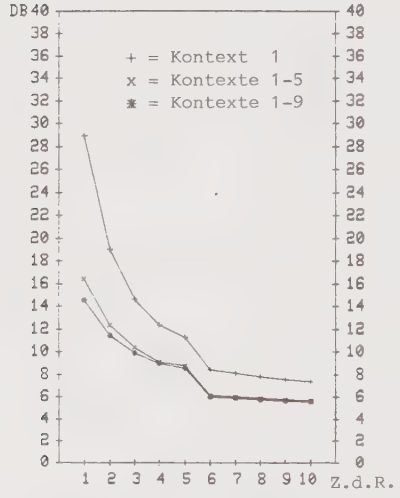


Abb. 5.32: ALEZ3.BRT

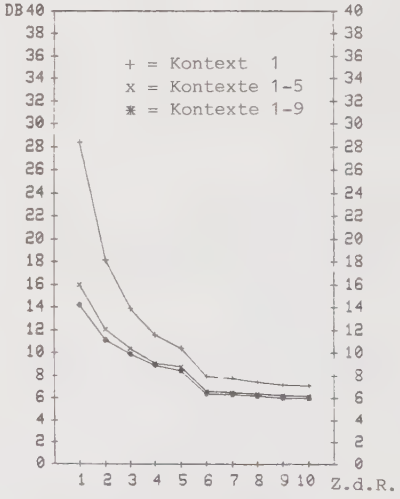


Abb. 5.33: ALEZ4.BRT

Abb. 5.30 - 5.33: Streubreiten für Einzelspektren, gemischte Kontexte, in Abhängigkeit von der Zahl der Realisierungen (Daten aus Tabellen 5.6 - 5.9, jeweils erste, fünfte und neunte Zeile)

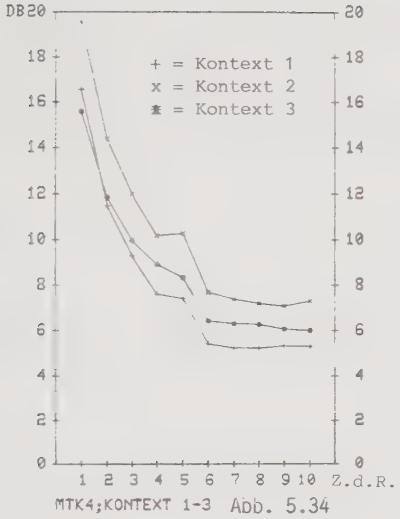


Abb. 5.34

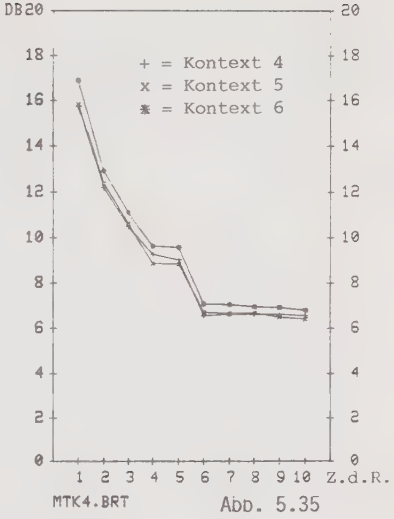


Abb. 5.35

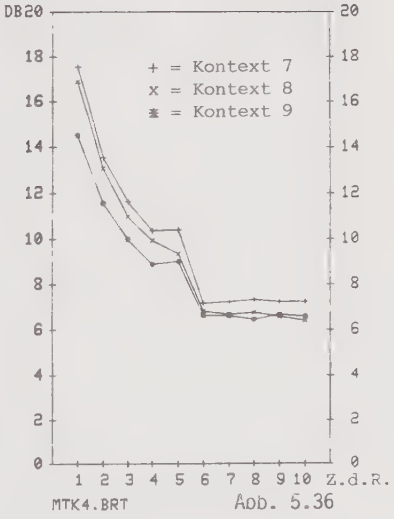


Abb. 5.36

Abb. 5.34 - 5.36:
Streubreiten für Mittelwertspektren nach Kontexten getrennt in Abhängigkeit von der Zahl der Realisierungen
(Daten aus Tabelle 5.13)

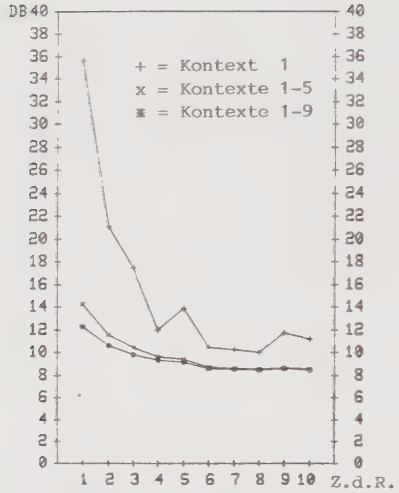


Abb. 5.37: ALMT1. BRT

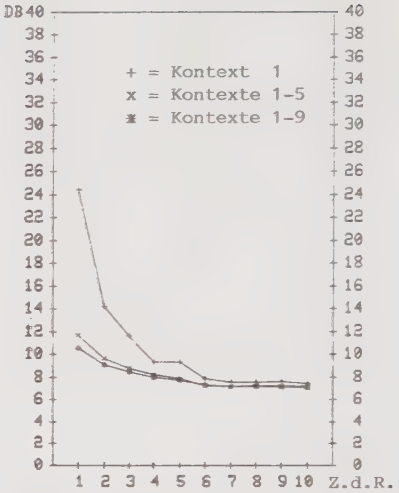


Abb. 5.38: ALMT2.BRT

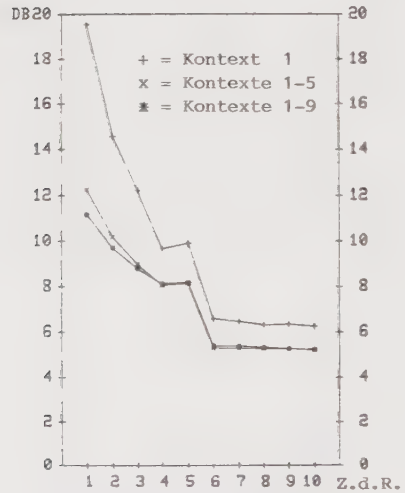


Abb. 5. 39: ALMT3.BRT

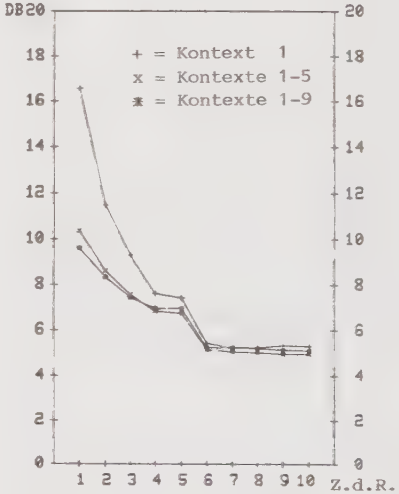


Abb. 5.40: ALMT4.BRT

Abb. 5.37 - 5.40: Streubreiten für Mittelwertspektren, gemischte Kontexte, in Abhängigkeit von der Zahl der Realisierungen (Daten aus Tabellen 5.14 - 5.17, jeweils erste, fünfte und neunte Zeile

6. KRIMINALTECHNISCHE ANWENDUNGEN

Für eine realistische Prüfung der Anwendbarkeit der bisher gewonnenen Erkenntnisse werden in diesem Kapitel nach einer kurzen Übersicht praxisgerechte Messungen vorgestellt.

6.1 Überblick

Die Einsatzmöglichkeiten des in den vorhergehenden Kapiteln geschilderten Verfahrens und sein praktischer Ablauf sind in Abbildung 6.1 dargestellt. Das Sprachsignal aus der Tataufzeichnung wird digitalisiert und auf Datenträger abgelegt. Als erster Schritt müssen die /f/-Realisierungen aus dem Tatgespräch segmentiert werden. Für die Segmentation kommt es nun darauf an, ob für die Ermittlung der Eigenschaften des Übertragungskanals die kontextabhängige oder die kontextunabhängige Vorgehensweise gewählt wird. Beide Methoden sind grundsätzlich möglich, da die Art der Aufarbeitung des zur Verfügung stehenden Hintergrundmaterials in Dateien mit Spektren aus /f/-Lauten einzelner Kontextkategorien und anderen mit gemischten Kontexten erfolgte. Die kontextabhängige Vorgehensweise bedeutet, daß nur diejenigen /f/-Realisierungen aus dem Tatmaterial segmentiert werden, die zu einer bestimmten, bereits untersuchten und für geeignet befundenen Kontextkategorie gehören. Bei der kontextunabhängigen Verfahrensweise können dagegen die /f/-Realisierungen aus verschiedenen Kontexten gemischt werden. Grundsätzlich sollten dabei nur solche /f/-Laute verwendet werden, die auch im Hintergrundmaterial vorhanden sind. Wie die bisherigen Untersuchungen zeigen, wird sich allerdings kein großer Fehler bei Hinzunahme von /f/-Spektren aus anderen Kontexten ergeben. Dies kann erforderlich werden, wenn das Tatgespräch nur kurz ist und insgesamt nur wenige /f/-Laute vorliegen. Ein für die forensische Praxis vermutlich entscheidender Vorteil liegt bei der kontextunabhängigen Messung, da hier auch bei kürzeren Texten schnell eine genügende Anzahl von /f/-Realisierungen vorhanden ist.

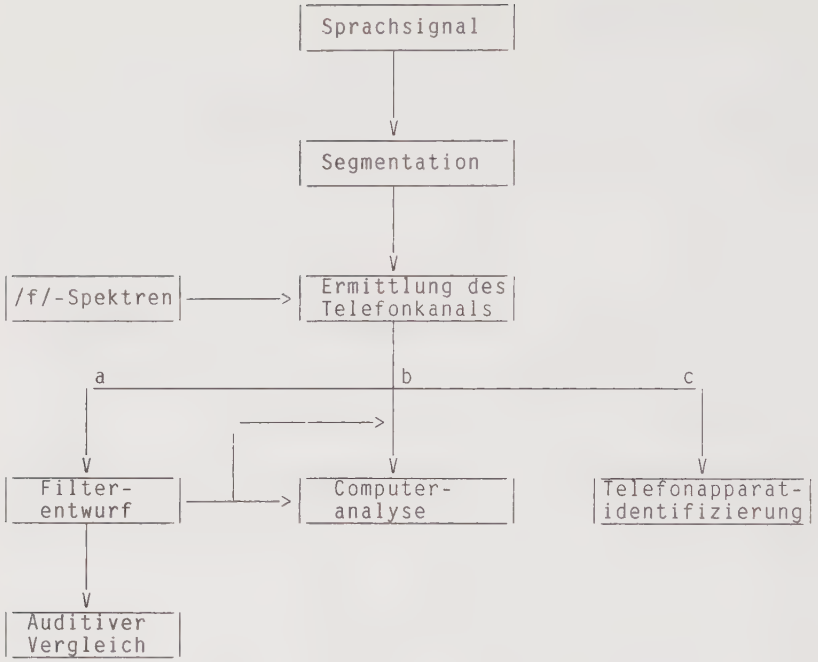


Abb. 6.1: Einsatzmöglichkeiten des beschriebenen Verfahrens im kriminaltechnischen Bereich

Nach Berechnung der Kanaleigenschaften können diese

a) für die Aufbereitung des Sprachmaterials für den auditiven Stimmenvergleich benutzt werden, (Ast a in Abbildung 6.1)

b) direkt zu Korrekturen bei spektralanalytischen oder sonstigen Computeranalysen dienen. Auch das ent- bzw. verzerrte Sprachsignal kann einer solchen Verarbeitung zugeführt werden. (Ast b in Abbildung 6.1)

c) unter bestimmten Voraussetzungen zur Telefonapparat-identifizierung benutzt werden. (Ast c in Abbildung 6.1)

In den folgenden Abschnitten werden die verschiedenen Anwendungsmöglichkeiten mit Ausnahme der Computeranalyse, die zugehörigen Messungen und das benutzte Datenmaterial, bestehend aus zum Vergleich ausgemessenen Telefonkanälen und entsprechenden Sprachaufzeichnungen im einzelnen vorgestellt.

6.2 Datensammlung

Die hier durchzuführenden Untersuchungen erfordern grundsätzlich zwei Arten von Daten: über Telefon übertragene Sprachproben, aus denen /f/-Spektren und daraus wiederum die Kanaleigenschaften berechnet werden und zusätzlich zum Vergleich die auf andere Weise vermessenen Übertragungseigenschaften der Kanäle.
Die größte Schwierigkeit bei der Sammlung des Datenmaterials war eine von der hier beschriebenen Methode unabhängige und zuverlässige Vermessung von Übertragungsstrecken, mit der diejenigen aus der /f/-Messung zur Kontrolle verglichen werden konnten. Dies war insofern schwierig, als die Abstrahlung eines bekannten Meßsignals mit gleicher Intensität im gesamten Frequenzbereich vor der Sprechkapsel eines Telefonapparates nicht unproblematisch ist. Die für solche Zwecke übliche Standardausrüstung wie Generator und künstlicher Mund, gekoppelt mit einem Pegelschreiber, konnte wegen der räumlichen Trennung von Sender und Empfänger in dieser Kombination nicht benutzt werden. Es wurden schließlich auf folgende Arten Daten gesammelt:

Datentyp 1:

In eine hochwertige Übertragungsstrecke wurden verschiedene Filter mit bekannten Eigenschaften eingebaut. Der Sprecher sprach direkt in ein Mikrophon (im Gegensatz zur Tonbandwiedergabe) und die gefilterte Sprache wurde mit einem Tonbandgerät aufgezeichnet. Diese Datensammlung sollte Messungen ermöglichen, die mit Sicherheit frei von nichtlinearen Verzerrungen sind. Zur Ermittlung der Übertragungsfunktion für den Vergleich wurde von einem

Generator erzeugtes weißes Rauschen von einem künstlichen Mund abgestrahlt. Dieser besteht aus einem Lautsprecher, an dessen Rand sich in definiertem Abstand und Winkel ein Mikrophon befindet. Mit einem Signalgenerator wird normalerweise der interessierende Frequenzbereich stetig durchfahren und die Intensität der abgestrahlten Energie bei sich verändernden Frequenzen über das eingebaute Mikrophon gemessen und auf einen festen Wert korrigiert. Mit dem Generator ist in der Regel ein Pegelschreiber gekoppelt, der die Meßwerte automatisch in einem Diagramm den erzeugten Frequenzen zuordnet. Für den Datentyp 1 (und Datentyp 2) wurde der künstliche Mund zur Abstrahlung des Rauschsignals benutzt und das Spektrum des vom eingebauten Mikrophon aufgenommenen Signals als Eichgröße verwendet. Auf diese Weise sollte die durch den Lautsprecher des Mundes selbst verursachte Frequenzverzerrung berücksichtigt werden. Dieses Signal stellt unter den gegebenen Umständen die bestmögliche Approximation dar, entspricht aber nicht exakt dem tatsächlich vorhandenen Schallfeld. Dies ist bei der späteren Beurteilung der Meßergebnisse zu berücksichtigen (siehe z.B. Abschnitt 6.2). Abbildung 6.2 zeigt das aufgenommene Spektrum, wenn weißes Rauschen (0-20 kHz) auf den Lautsprecher des Mundes gegeben wird. Davon abweichende Strukturen im empfangenen Signalspektrum sind von der Übertragungsstrecke verursacht worden.

Datentyp 2:

Ein weiterer Datensatz wurde von verschiedenen gewöhnlichen Fernsprechapparaten (Hausapparate) über Orts- und Fernnetze aufgenommen. Die Sprecher benutzten das Telefon in der ihnen gewohnten Weise. Die Messung der Übertragungsstrecke erfolgte wie bei Datentyp 1. Die dazu notwendigen Geräte wurden zu den sendenden Apparaten transportiert. Die Abnahme des Signals am empfangenen Apparat erfolgte meist mit einem normalen Telefonadapter, in einigen Fällen auch durch Abgriff an der Hörermuschel. Letzteres führte zu einer besseren Anpassung und damit zu einem größeren Rauschabstand. Hierauf wird in der Diskussion soweit nötig hingewiesen.

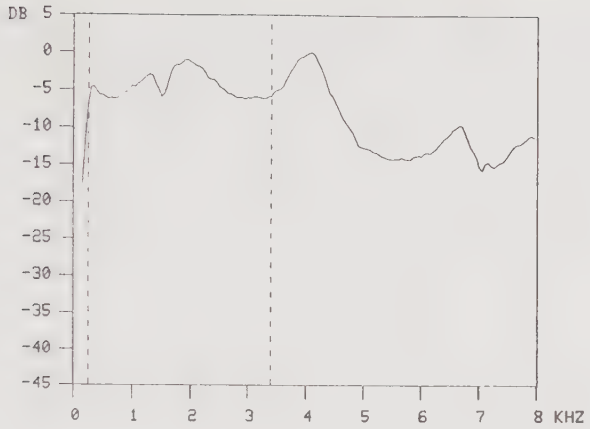


Abb. 6.2: Spektrum, aufgenommen vom Mikrophon des künstlichen Mundes bei weißem Rauschen

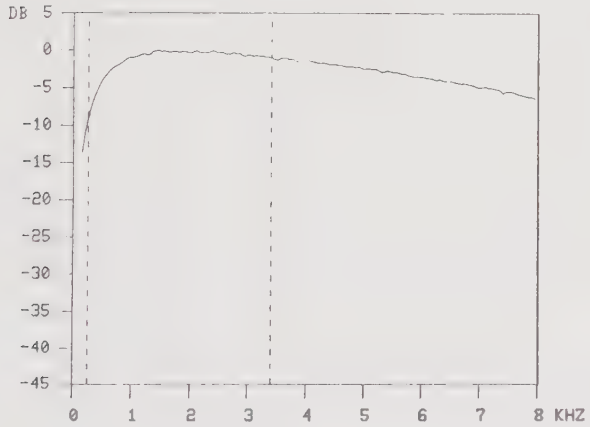


Abb. 6.3: Spektrum des Rauschens vom Tonband für Messungen am Telefonkanal

Datentyp 3:
Zur Messung an Münzfernsprechern wurde der Testsatz und anschließend weißes Rauschen in den Telefonzellen von einem Tonband über eine Kopfhörermuschel wiedergegeben, die mit einem Schaumstoffpolster manuell auf den Mikrophonteil des Telefonhörers gedrückt wurde. Die Messung wurde aus verschiedenen Telefonzellen innerhalb des Ortsnetzes Wiesbaden und aus einem benachbarten Ortsnetz (Taunusstein) vorgenommen. Ein Teil der Telefonzellen befand sich in geräuschvoller Umgebung, z.B. an einer Kreuzung von Hauptstraßen sowie an einer Fernstraße. Auf diese Weise sollten praxisgerechte erschwerte Bedingungen insbesondere auch für die Segmentation geschaffen werden. Von dem benachbarten Ortsnetz aus wurde weiterhin eine Leitung mit einem gewöhnlichen Hausapparat vermessen. Als Quellenspektrum für die Messung mit weißem Rauschen wurde das Spektrum gewählt, wie es vom Tonband direkt wiedergegeben wurde (Abbildung 6.3). Veränderungen durch die Kopfhörermuschel bzw. durch das mehr oder weniger fest und nicht immer konstant zusammengepreßte Kunststoffpolster wurden nicht berücksichtigt.

6.3 Messungen an simulierten Übertragungskanälen mit Filtern

In diesem Abschnitt soll an Beispielen untersucht werden, wie gut sich Eigenschaften einer bekannten und nachprüfbaren Telefonleitung nachträglich aus dem übertragenen Sprachsignal wiedergewinnen lassen, wenn keine nichtlinearen Verzerrungen vorliegen (nach Datentyp 1). Aus diesem Grund wurden zwei Übertragungswege benutzt, in die je ein Filter eingebaut wurde, dessen Frequenzcharakteristik leicht ausgemessen werden konnte. (Abbildungen 6.4 und 6.5). Über jedes dieser Filter wurde von einer Testperson, die nicht mit einem der 10 Sprecher aus dem Untersuchungsmaterial identisch ist, der in Abschnitt 4.2.5 genannte Testsatz gesprochen und anschließend der gesamte Übertragungsweg einschließlich Mikrophon mit weißem Rauschen vermessen. Aus jedem

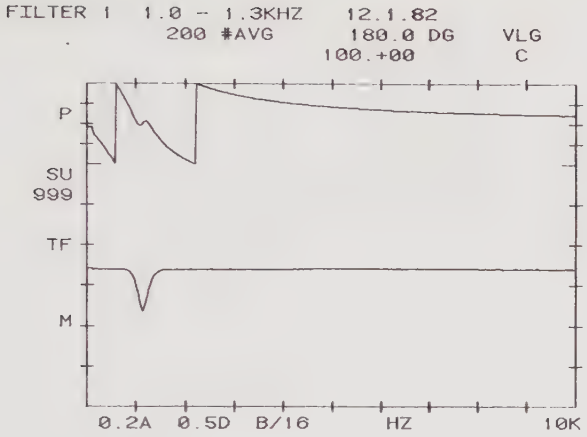


Abb. 6.4: Übertragungscharakteristik Filter 1

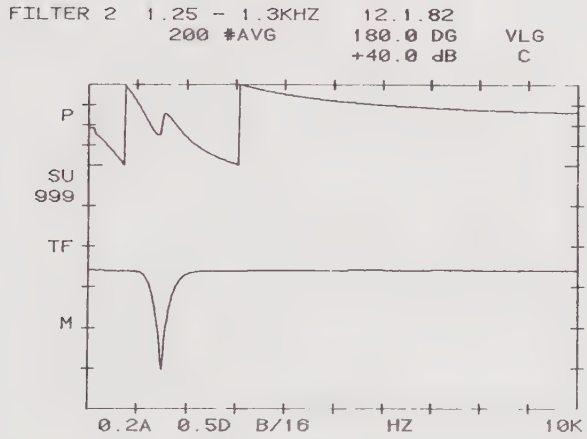


Abb. 6.5: Übertragungscharakteristik Filter 2

der beiden Sätze wurden dann die neun /f/-Laute segmentiert und für die Berechnung benutzt.

Die Abbildungen 6.6 und 6.7 zeigen die Ergebnisse für diese "künstlichen" Übertragungsstrecken. Die mittlere durchgezogene Kurve ist das durch /f/-Messung gewonnene Spektrum; die ebenfalls durchgezogenen äußeren Kurven zeigen den Streubereich, der sich bei Benutzung von neun /f/-Lauten (extremwertbereinigt, siehe Abschnitt 5.4)) auf der Basis des untersuchten Hintergrundmaterials ergibt. Die gestrichelte Linie ist das Ergebnis der Messung mit weißem Rauschen. Die Abweichungen dieser Kurve von den Abbildungen in 6.4 und 6.5 stammen im wesentlichen vom benutzten Mikrophon, können aber zum Teil auch auf die in Abschnitt 6.1 genannten methodischen Fehlermöglichkeiten im Meßverfahren beruhen.

6.4 Kontextunabhängige Messung an echten Telefonkanälen

Um zu realistischeren Testergebnissen für die kriminaltechnischen Anwendungen zu kommen, insbesondere auch um die Auswirkungen möglicher nichtlinearer Verzerrungen mit erfassen zu können, wurden echte Telefonverbindungen mittels /f/-Spektren vermessen. Dabei sollte sowohl die Möglichkeit der kontextunabhängigen wie auch abhängigen Messung geprüft werden. In diesem Abschnitt werden die je nach Menge des Materials für die Kriminaltechnik besonders bedeutsamen kontextunabhängigen Messungen vorgestellt, d.h. solche, bei denen /f/-Laute aus verschiedenen lautlichen Umgebungen verwendet werden. Das als Eichsignal zu verwendende Hintergrundspektrum soll dabei aus einer entsprechenden Vielfalt verschiedener Kontexte und aus einer möglichst großen Zahl von Sprechern und Realisationen hervorgegangen sein. Dabei werden in diesen Experimenten u.a. auch /f/-Laute mit einbezogen, die nicht im Hintergrundmaterial enthalten sind. Diese Voraussetzungen bietet im Rahmen dieser Untersuchung das Spektrum in Abbildung 5.12 (Mittelung über alle Sprecher, Kontexte und Reali-

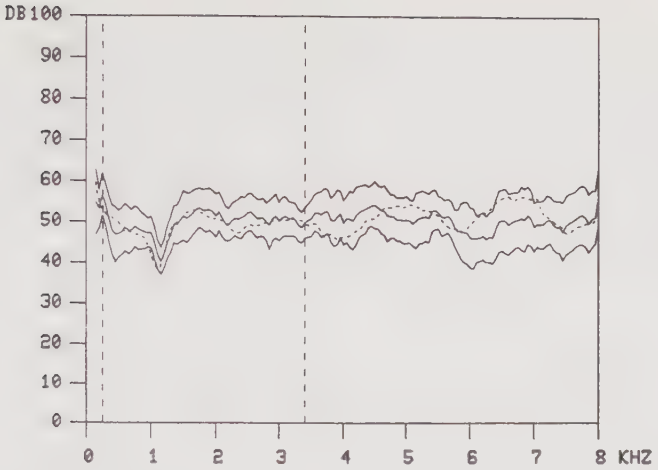


Abb. 6.6: Messungen mit Filter 1. Die mittlere durchgezogene Linie ist der mittels /f/-Spektren bestimmte Übertragungsweg. Die gestrichelte Kurve zeigt die Messung mit weißem Rauschen

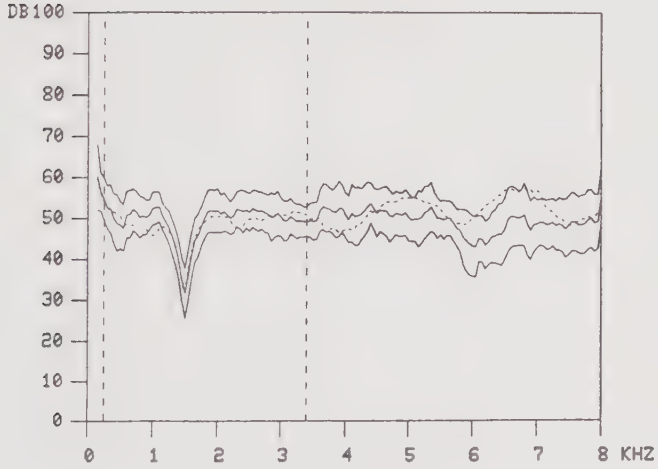


Abb. 6.7: Messungen mit Filter 2. Bezeichnungen wie in Abbildung 6.6

sationen), das für alle Messungen dieses Abschnittes verwendet wird. Untersuchungen wurden durchgeführt an Orts- und Fernleitungen unter Benutzung von normalen Haustelefonapparaten mit Kohlemikrophonen sowie aus öffentlichen Münzfernsprechern mit dynamischen Mikrophonen (Datentypen 2 und 3). Die Ergebnisse sind in den Abbildungen 6.8 bis 6.17 dargestellt und werden im folgenden diskutiert.

6.4.1 Ortsleitungen mit gewöhnlichen Fernsprechapparaten

Die Abbildungen 6.8 und 6.9 zeigen Ergebnisse für eine Ortsleitung, für die der sendende Apparat zunächst mit einer (hier fabrikneuen) Kohlekapsel, dann mit einer Kapsel mit eingebautem Verstärker ausgerüstet war. Die Telefonleitung wurde beibehalten. Das Testmaterial war vom Datentyp 2, Abnahme durch Adapter. Durch Vergleich der beiden getrichelten Meßkurven ist zu erkennen, daß die Übertragungscharakteristiken sich in beiden Telefonverbindungen deutlich unterscheiden. Dieser maßgebliche Einfluß der Sprechkapsel sollte eine Identifizierung des betreffenden Telefonapparates ermöglichen, wie sie in Abschnitt 6.6 beschrieben wird. Wie leicht zu sehen ist, wird durch die Verstärkerkapsel eine Frequenzbeschneidung durchgeführt. Das Sprachmaterial für diese Messungen bestand aus demselben Testsatz, der bei der Untersuchung der /f/-Laute verwendet wurde, so daß in beiden Fällen dieselben Kontextverhältnisse vorliegen. Da dies in der kriminaltechnischen Wirklichkeit wohl nur in den Fällen zutrifft, in denen sehr umfangreiche Aufzeichnungen zur Verfügung stehen oder der Verdächtige bereit ist, den Text aus dem Tatgespräch zu wiederholen, wurden weitere Versuche mit anderem Sprachmaterial durchgeführt. Das Testmaterial bestand aus folgendem - im Rahmen anderer Untersuchungen bereits benutztem - "Drohanruf":

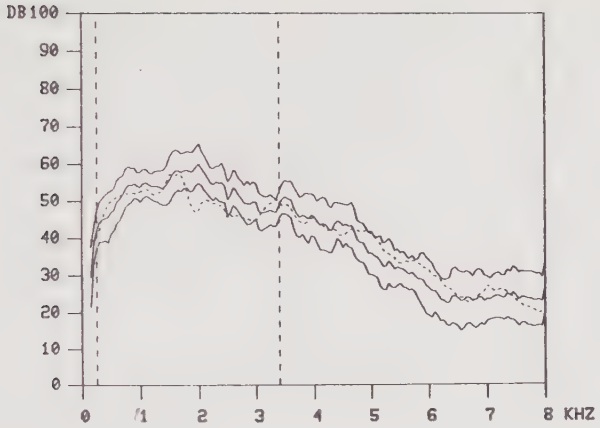


Abb. 6.8: Messungen an Ortsleitung 2, Kohlemikrophon, neun verschiedene /f/-Laute,
----- Messung mit weißem Rauschen

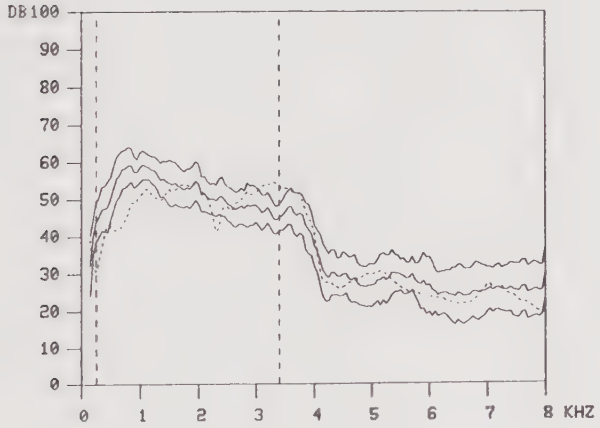


Abb. 6.9: Messungen an Ortsleitung 2, Verstärkerkapsel, neun verschiedene /f/-Laute
----- Messung mit weißem Rauschen

"Guten Tag, hier ist Meier, es geht um folgendes: Wir haben Ihre kleine Tochter Ramona nach der Schule in unsere Obhut genommen. Wenn Sie nicht wollen, daß ihr etwas passiert, dann hören Sie jetzt mal gut zu: Besorgen Sie sich vierzigtausend Mark. Packen Sie das Geld in einen schwarzen Koffer. Sie spielen dann selbst den Boten. Fahren Sie mit Ihrem Wagen in Richtung Scheef. Bei dem Schuppen an der Ausfahrt Camberg bleiben Sie stehen. Und noch etwas: Lassen Sie die Polizei aus dem Spiel. Wenn Sie den Schuppen erreicht haben, stellen Sie den Motor ab und bleiben Sie im Wagen sitzen. Sie hören dann von uns. Wenn Sie glauben, Sie könnten quer schießen, dann liegen Sie schief, dann kriegen Sie keinen Fuß mehr auf die Erde. Wir werden Sie auf gar keinen Fall schonen und das wäre doch sehr schade."

Insgesamt sind elf /f/-Realisierungen (unterstrichen) in dem Material enthalten, die auch alle ausgewertet wurden. Die Abbildungen 6.10 und 6.11 zeigen die Ergebnisse für zwei verschiedene Ortsleitungen. In beiden Fällen ist eine zufriedenstellende Übereinstimmung beider Kurvenverläufe festzustellen. Dies bedeutet, daß das zur Messung benutzte Hintergrundspektrum (aus Abbildung 5.12) aus genügend vielen bezüglich ihres Kontextes unterschiedlichen /f/Lauten entstanden und damit generell als Meßspektrum für die Berechnung der Übertragungsstrecke mit /f/-Spektren aus gemischten Kontexten einsetzbar ist. Streubreiten wurden hier nicht angegeben, da die aus 9 /f/-Realisierungen ermittelte Datei ALMT91.GRZ die Verhältnisse aus den in Abschnitt 5.4 bereits geschilderten Gründen für 11 /f/-Laute unrealistisch schlecht wiedergibt. Im wesentlichen dürften aber ungefähr dieselben Streubreiten (etwa +/- 5 dB) zutreffen.

6.4.2 Ortsleitungen mit öffentlichen Münzfernsprechern

Wie beschrieben wurden mehrere Ortsleitungen mit öffentlichen Münzfernsprechern aus dem Ortsnetze Wiesbaden vermessen (Datentyp 3). Das Signal am Empfänger wurde mit einem Adapter abgenommen. Die Abbildungen 6.12 und 6.13 zeigen einige Ergebnisse. In beiden Beispielen ist eine gute Rekonstruktion der Übertragungsstrecke festzustellen. Der flachere (und wirklichkeitsgerechtere) Verlauf der mit Rauschen gemessenen Kurven oberhalb von 4.5 kHz ist darauf zurückzuführen, daß der Rauschabstand des Signals für die /f/-Laute bei Abnahme durch ein Adapter nicht groß genug war. In allen Messungen mit gutem Rauschabstand (z.B. durch optimalem Signalabgriff an der Hörmuschel, war dieser unterschiedliche Kurvenverlauf nicht zu beobachten.

6.4.3 Leitungen aus dem Nahbereich

Von einem benachbarten Ortsnetz im Nahbereich Wiesbaden wurden ebenfalls von verschiedenen Telefonzellen (Datentyp 3) sowie von einem Hausapparat (Datentyp 2) Sprachproben übertragen und die Kanäle vermessen. Die Ergebnisse sind in den Abbildungen 6.14 und 6.15 dargestellt. In Abbildung 6.15 ist wieder der zu flache Verlauf der durch /f/-Messung gewonnenen Kurve aufgrund der schlechten Anpassung des Adapters zu beobachten.

6.4.4 Leitungen aus der Fernebene

Für Messungen an Fernverbindungen wurde das Testmaterial und weißes Rauschen (Datentyp 2) unter Verwendung des künstlichen Mundes über zwei getrennt gewählte Leitungen Wiesbaden-Hannover-Wiesbaden übertragen. Die Abbildungen 6.16 und 6.17 zeigen die Ergebnise der Berechnung. Für die in Abbildung 6.16 dargestellte Messung wurde eine Adapterabnahme gewählt. Die /f/-Laute ent-

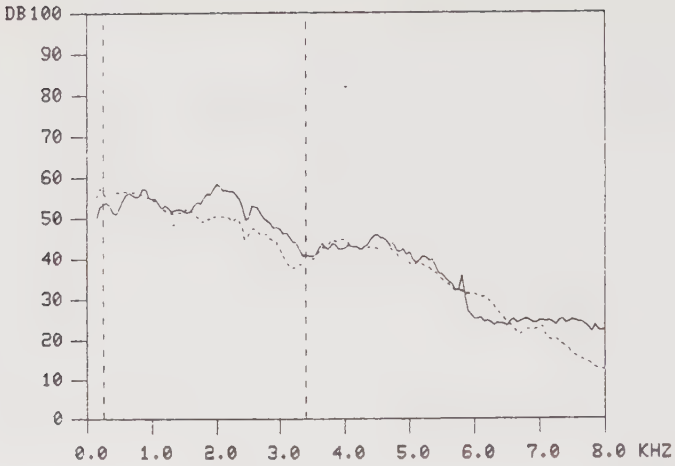


Abb. 6.10: Messungen an Ortsleitung 3, Kohlemikrophon, elf verschiedene /f/-Laute,
----- Messung mit weißem Rauschen

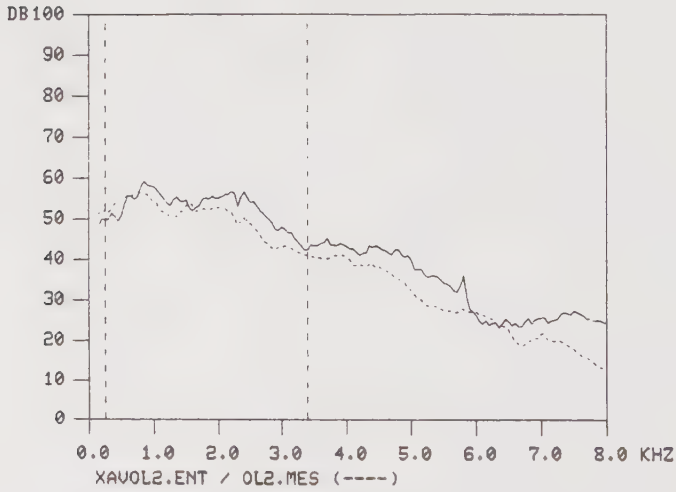


Abb. 6.11: Messungen an Ortsleitung 4, Kohlemikrophon, elf verschiedene /f/-Laute,
----- Messung mit weißem Rauschen

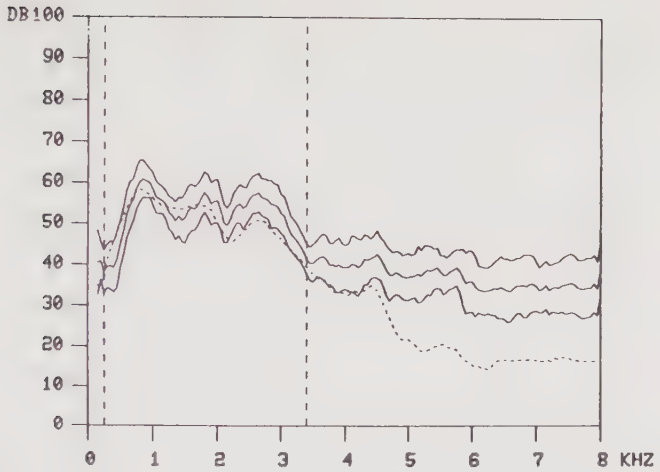


Abb. 6.12: Messungen an einer Ortsleitung mit Telefonzelle, neun verschiedene /f/-Laute,
----- Messung mit weißem Rauschen

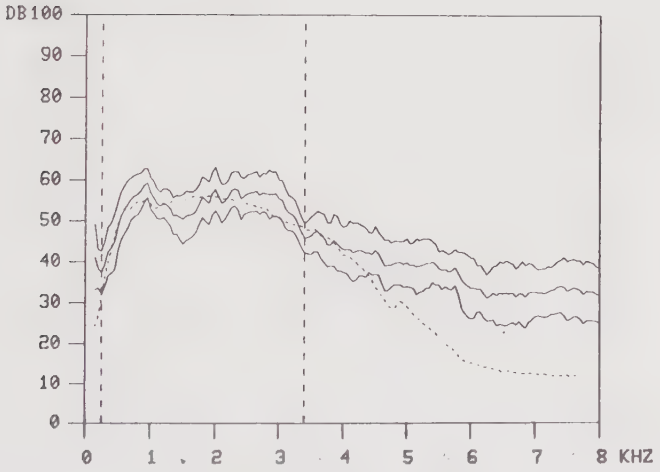


Abb. 6.13: Messungen an einer Ortsleitung mit Telefonzelle, neun verschiedene /f/-Laute,
----- Messung mit weißem Rauschen

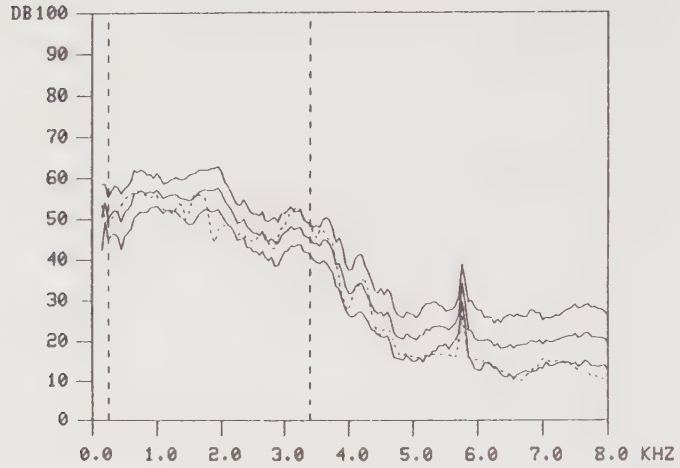


Abb. 6.14: Messungen im Nahbereich, Hausapparat,
neun verschiedene /f/-Laute,
----- Messung mit weißem Rauschen

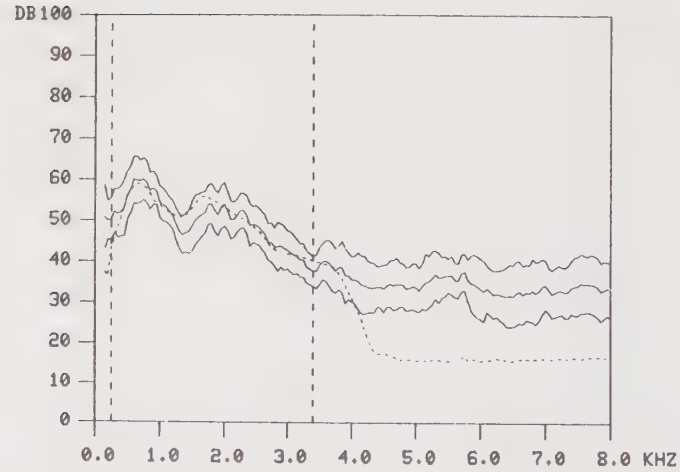


Abb. 6.15: Messungen im Nahbereich, Telefonzelle,
neun verschiedene /f/-Laute
----- Messung mit weißem Rauschen

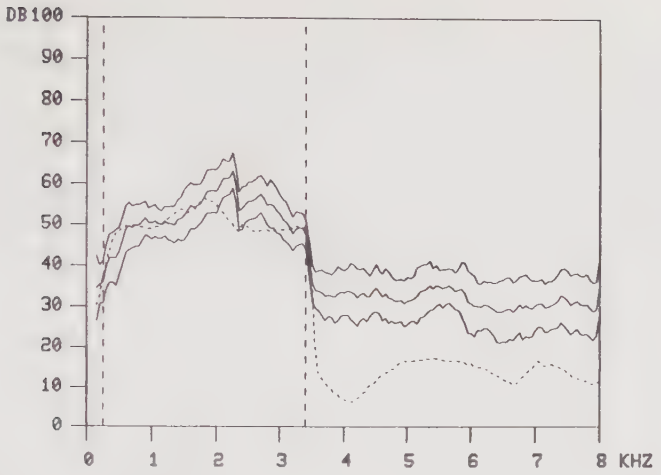


Abb. 6.16: Messungen an einer Fernleitung, Kohlemikrophon, neun verschiedene /f/-Laute,
----- Messung mit weißem Rauschen

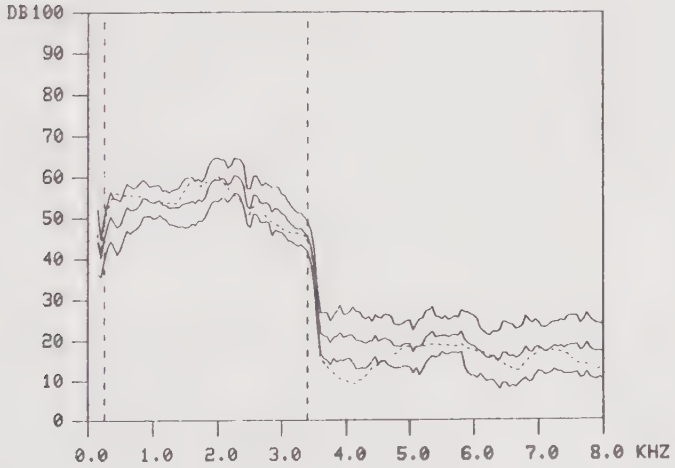


Abb. 6.17: Messungen an einer Fernleitung, Kohlemikrophon, elf verschiedene /f/-Laute,
----- Messung mit weißem Rauschen

stammen dem beschriebenen Testsatz. In Abbildung 6.17 wurde dagegen das Signal an der Hörmuschel abgegriffen und der Text aus dem "Drohanruf" verwendet. Als Streubreitenangabe wurde hier die Datei ALMT91.GRZ eingezeichnet, die wie erwähnt die Verhältnisse für 11 Spektren eher zu schlecht wiedergibt. In beiden Messungen ist eine recht gute Rekonstruktion der Übertragungsstrecke festzustellen, so wird u.a. die Frequenzbeschneidung bei 3.4 kHz gut wiedergegeben.

6.5 Kontextabhängige Messungen an echten Telefonkanälen

Wie bereits erläutert sind die Ausgangsdaten so aufbereitet worden, daß auch die Möglichkeit zur kontextabhängigen Vermessung besteht. Der Übersichtlichkeit wegen werden hier nur Meßbeispiele anhand der Kontexte 1 (/f/ in /Vater/) und des Kontextes 4 (/f/ in /fach/) vorgestellt. Diese /f/-Laute wurden gewählt, weil einerseits in beiden Fällen die Segmentation verhältnismäßig problemlos war und sich andererseits geringe Streubreiten ergeben. Für Kontext 1 werden bei zehn /f/-Realisierungen für den Telefonkanal extremwertbereinigt +/- 2.65 dB und für Kontext 4 unter den gleichen Bedingungen +/- 3.28 dB erreicht. Das Sprachmaterial wurde wie in Datentyp 2 übertragen, allerdings sprachen die Sprecher dabei zehn mal nur die benötigten Wörter (z.B. /Vater/ und /vielfach/. Die Abbildungen 6.18 und 6.19 zeigen die Ergebnisse für die Messung an einer identischen Ortsleitung mit Meßdaten der genannten Kontexte.

Zum Vergleich der Genauigkeit von kontextabhängiger und kontextunabhängigen Vermessung wurden eine Fernleitung und eine Ortsleitung auf beide Arten bestimmt und in den Abbildungen 6.20 und 6.21 eingetragen. In Abbildung 6.20 wurde für die Messung mit gemischten Kontexten der Text aus dem "Drohanruf" und für die kontextabhängige Messung die /f/-Realisierungen in /vielfach/ (Kontext 4) gewählt. Den Ergebnissen in Abbildung 6.21 liegen der Testsatz bzw. der Kontext 1 zugrunde. Wie leicht zu sehen ist, sind in beiden Messungen die Kurvenverläufe nahezu gleich.

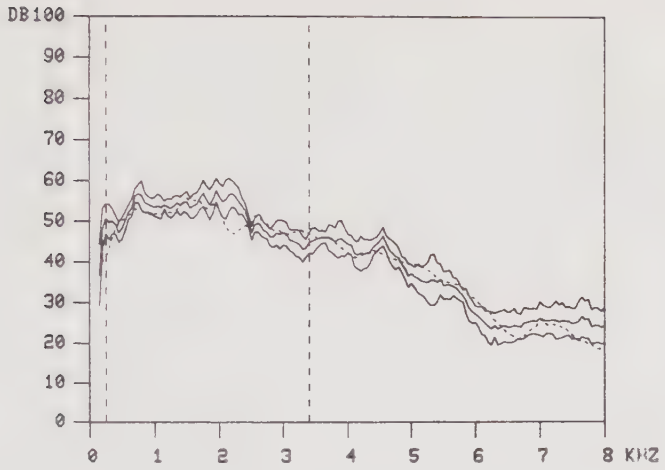


Abb. 6.18: Messungen an einer Ortsleitung mit Kontext 1,
10 /f/-Realisierungen aus /Vater/,
----- Messung mit weißem Rauschen

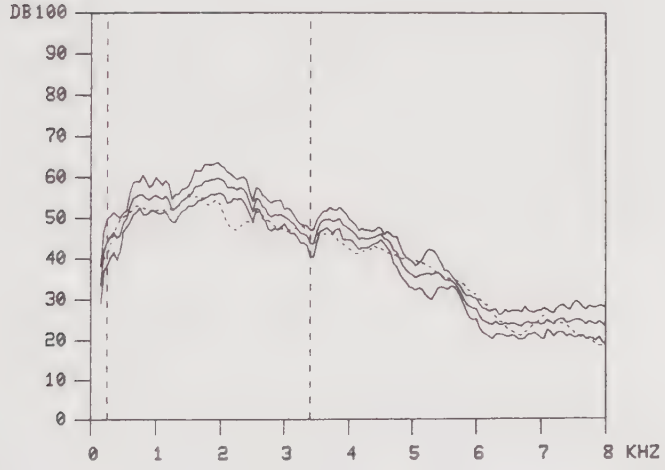


Abb. 6.19: Messungen an einer Ortsleitung mit Kontext 4,
10 /f/-Realisierungen aus /vielfach/,
----- Messung mit weißem Rauschen

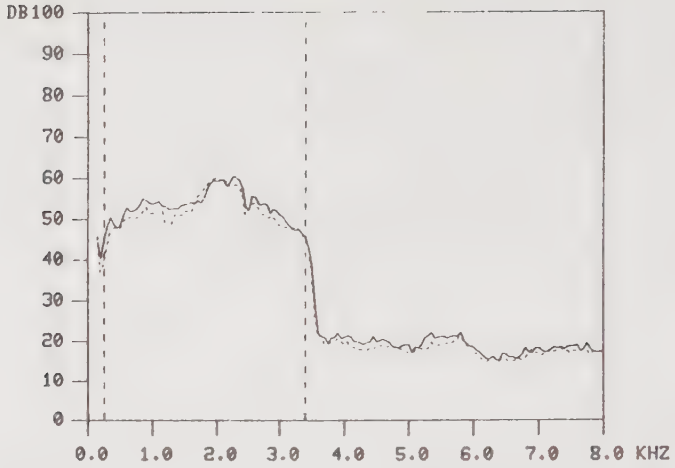


Abb. 6.20: Vergleich von kontextunabhängiger und kontextabhängiger Messung an einer Fernleitung
----- Messung mit Kontext 4 (vielfach)
——— kontextunabhängige Messung ("Drohanruf")

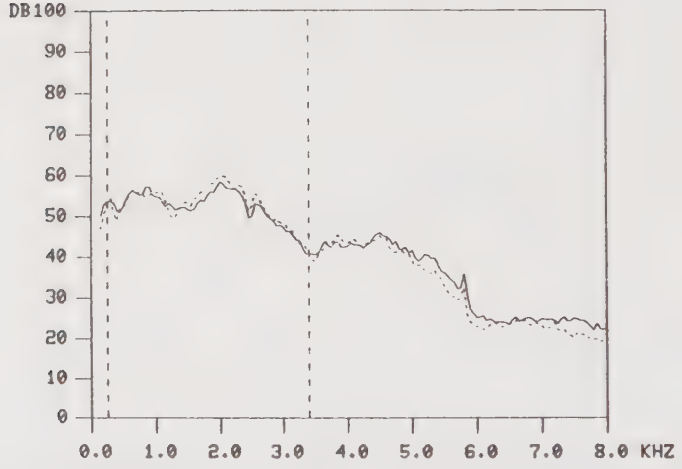


Abb. 6.21: Vergleich von kontextunabhängiger und kontextabhängiger Messung an einer Ortsleitung
----- Messung mit Kontext 4 (vielfach)
——— kontextunabhängige Messung (Testsatz)

6.6 Identifizierung von Fernsprechapparaten

Wie in Kapitel 1 bereits geschildert, kann die Identifizierung eines Telefonapparates, von dem aus ein Tatgespräch geführt worden ist, aus einer Menge von vorgegebenen möglichen Apparaten für polizeiliche Zwecke von großer Bedeutung sein. In einem solchen Fall könnten die infrage kommenden Leitungen - z.B. mit weißem Rauschen - vermessen und die sich ergebenden Übertragungsfunktionen mit den aus dem Tatgespräch durch /f/-Spektren gewonnenen verglichen werden. In den Abbildungen 6.22 und 6.23 sind für eine Ortsleitung mit einem gewöhnlichen Telefonapparat sowie für eine andere Ortsleitung mit Münzfernsprecher die mit Rauschen gemessenen Übertragungseigenschaften der betreffenden und weiterer Telefonverbindungen sowie die aus der /f/-Messung bestimmte Kurve (gestrichelt) dargestellt.

Abbildung 6.22 ist ein Beispiel dafür, daß nicht immer eine Zuordnung erfolgen kann. Im Bereich zwischen 1.5 und 3.0 kHz haben die mit OL1 bezeichneten Kurven die beste Übereinstimmung, dafür zeigen von 0.5 bis 1.5 kHz OL1 und OL2 einen gleichartigen Verlauf. Eine Entscheidung ist - zumindest bei dem derzeitigen Wissensstand - nicht möglich. In Abbildung 6.23 dagegen ist eine Zuordnung leicht zu treffen. Das infrage kommende Gespräch wurde vermutlich von Apparat 2 geführt.

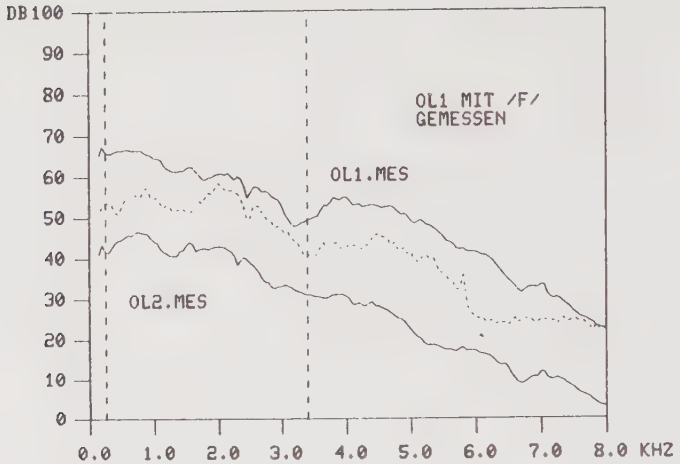


Abb. 6.22: Vergleich mehrerer gemessener mit einer durch /f/-Spektren bestimmten Übertragungsstrecke
----- Messung mit /f/-Spektren
(Hausapparate)

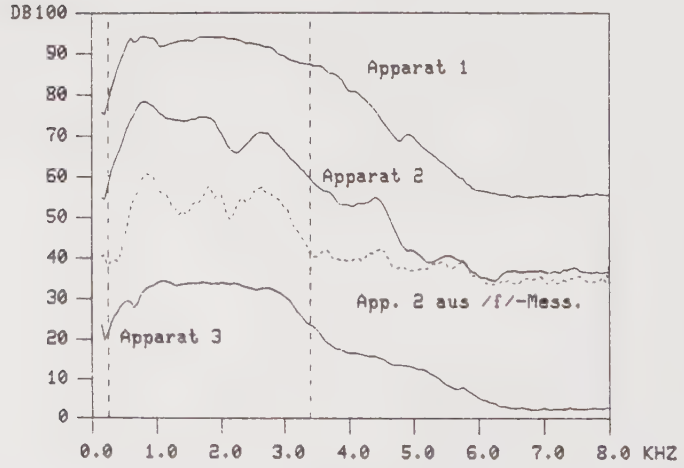


Abb. 6.23: Vergleich mehrerer gemessener mit einer durch /f/-Spektren bestimmten Übertragungsstrecke
----- Messung mit /f/-Spektren
(Telefonzellen)

6.7 Ent- und Verzerrung von Sprachaufzeichnungen für den Stimmenvergleich

Die Erfahrung zeigt, daß der typische Stimmenvergleich in der Kriminaltechnik aus dem Vergleich einer Aufzeichnung mit Telefonqualität (die Tataufnahme) und einer Aufnahme mit erheblich besseren Bedingungen, oft z.B. ohne bandbegrenzende Übertragungsstrecke, besteht. Der auditive Vergleich wird durch diese unterschiedlichen Qualitäten teilweise beeinträchtigt, da die spektralen Eigenschaften der Stimme erheblich durch die Übertragungseigenschaften der Telefonstrecke verändert werden können. Hier liegt nun nahe, das beschriebene Verfahren dazu einzusetzen, bessere Voraussetzungen für den auditiven Vergleich zu schaffen Ast a in Abbildung 6.1). Dazu bieten sich zwei Möglichkeiten:

- Nach Ermittlung der Übertragungseigenschaften des Telefonkanals wird dem hochwertigen Signal dieselbe Charakteristik aufgeprägt wie der Tataufnahme. Die Vergleichsaufnahme wird dann im nachhinein sozusagen durch dieselbe Telefonleitung gesendet wie das Tatgespräch (Abbildung 6.24(b)).

- Die Tataufnahme wird entzerrt, d.h. der Einfluß der Übertragungsstrecke wird nachträglich aus dem Telefonkanal entfernt. Bei Ferngesprächen muß zum Vergleich die hochwertige Aufnahme dennoch band begrenzt werden, da die Tataufnahme oberhalb von 3.4 kHz nicht wieder rekonstruiert werden kann (Abbildung 6.24(b)).

Ent- und Verzerrung des Sprachsignals erfordern eine aufwendige Software, die die Möglichkeit zum bequemen Filterentwurf sowie einer effektiven Filterung bestehen muß.*)

*)In Zusammenarbeit mit Herrn Dr. Borys, jetzt Universität Kassel, der die Software für den Filterentwurf fertigte, wurde mit Untersuchungen zu diesem Einsatzbereich begonnen. Die ersten Ergebnisse waren vielversprechend.

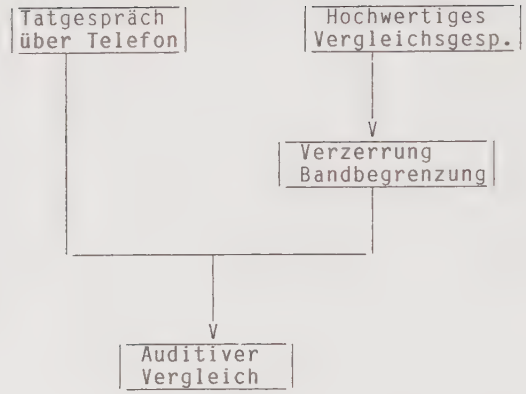


Abb. 6.24(a): Verzerrung der hochwertigen Vergleichsaufnahme für den auditiven Stimmenvergleich

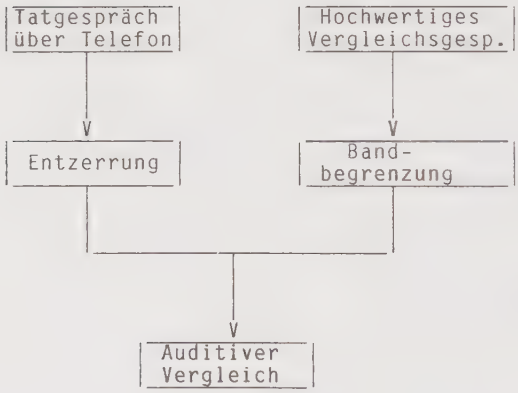


Abb. 6.24(b): Entzerrung des Tatgesprächs für den auditiven Stimmenvergleich

7. SCHLUSSFOLGERUNGEN UND AUSBLICK

In der vorliegenden Arbeit wird gezeigt, daß es möglich ist, die Eigenschaften einer Übertragungsstrecke allein aus dem übertragenen Sprachsignal durch Messung der Abweichung der Spektren von /f/-Realisierungen von entsprechenden bekannten Spektren bis zu einem gewissen Grad wiederzugewinnen. Dazu wurde die Spektralstruktur des /f/ eingehend untersucht. Ein solches Verfahren kann in der Kriminalistik erfolgreich auf dem Gebiet des Stimmenvergleiches sowie zur Identifizierung des Telefonapparates, von dem ein bestimmtes Gespräch geführt wurde, eingesetzt werden. Verbesserungen des Verfahrens können noch durch eine optimale Auswahl der Kontexte der verwendeten /f/-Laute erreicht werden. So wird sich wahrscheinlich ein Weglassen des außerordentlich anfälligen /f/ in Kontexten wie /schläft viel/ bei der Messung mit gemischten Kontexten positiv auswirken, es sei denn, die zum Vergleich herangezogenen /f/-Spektren aus dem Hintergrundmaterial entstammen denselben lautlichen Umgebungen wie im tatsächlichen Fallmaterial. Eine endgültige Bewertung über den Nutzen der hier entwickelten Methode kann allerdings nicht durch Testmessungen sondern nur durch Einsatz in der forensischen Praxis gewonnen werden.

8. LITERATURDOKUMENTATION

In der vorliegenden Arbeit wird nicht auf jede der hier aufgeführten Literaturstellen Bezug genommen. Es scheint jedoch sinnvoll, dem Leser wenigstens den wichtigsten Teil der umfangreichen Literatur zu dem außerordentlich komplexen Problem des forensischen Stimmenvergleiches und des hier behandelten speziellen Teilaspektes anzubieten. Dies gilt umsomehr, als aufgrund des interdisziplinären Charakters dieser Themen nur selten Arbeiten aus den verschiedenen Bereichen gemeinsam in Literaturverzeichnissen aufgeführt werden. Aus diesem Grund sind im folgenden insbesondere auch Veröffentlichungen berücksichtigt worden, die aus dem kriminaltechnischen oder forensischen Bereich stammen. Wegen der engen Verzahnung von Sprach- und Sprechererkennung sind auch einige Arbeiten aus dem Bereich der Spracherkennung, vorzugsweise solche mit Übersichtscharakter, aufgenommen worden. Weiterhin finden sich Literaturangaben zur digitalen Signalverarbeitung, insbesondere zur Behandlung stochastischer Signale.

AGARD (1982): Advanced Avionics and the Military Aircraft Man/Machine Interface. AGARD Conference Preprint No. 329.

AKIAYAMA, K. (1981): An introduction to the study of voice prints. Bulletin of the Phonetic Society of Japan, 168, 1-3.

ATAL, B.S. (1972): Text-independent speaker recognition. Paper presented at the 83th Meeting of the Acoustical Society of America, April 18-21, Buffalo, New York.

ATAL, B.S. (1972): Automatic speaker recognition based on pitch contours. Journal of the Acoustical Society of America, 52, 1687-1697.

ATAL, B.S. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. Journal of the Acoustical Society of America,

ATAL, B.S. (1964): Automatic recognition of speakers from their voices. Proc. IEEE, 460-475.

ATAL, B.S., HANAUER, S.L. (1971): Speech analysis and synthesis by linear prediction of the speech wave. Journal of the Acoustical Society of America, 50, 637-655.

BARTHOLOMEUS, B. (1973): Voice identification by nursery school children. Canadian Journal of Psychology, 27/4, 464-472

BASZTURA, C. (1981): Automatic speaker recognition by zero-crossing analysis of the speech signal. In: Speech Analysis and Synthesis 5, 9-40.

BATCHELLOR, M.P. (1981): Investigation of Parameters Affecting Voice Recognition Systems in C3 Systems. Naval Postgraduate School, Monterey, CA 93940, Master's Thesis.

BECKER, R.W. (1966): Characteristics that determine speaker recognition. Electronic System Division, ESD-TR-66-636, 1-66.

BEEK, B., NEUBERG, E.P., HODGE, D.C. (1977): An assessment of the technology of automatic speech recognition for military applications. In: IEEE-ASSP 25, 310-322.

BEEK, B. (1982): Overview of state-of-the-art, R & D NATO Activities, and possible Applications - Voice Processing Technology. AGARD Conference Preprint Nr. 329, (9-1 bis 9-15) Fachinformationszentrum Energie, Physik, Mathematik GmbH, 7514 Eggenstein.

BIDDULPH, R. (1952): Automatic recognition of spoken digits. Journal of the Acousticcal Society of America, 24,6, 637-642.

BIERFERT, H., VAN DER GIET G., GLAVE R.D., LANCE, D., REINHARD, C. (1977): Das DAWID Spracherkennungssystem: Stand und Anwendungen. DFG-Kolloquium, Muggendorf 29.-30. September.

BIERFERT, H., KIRSTEIN, M., LANCE D. (1980): Some aspects of evaluating the performance of a speech recognition system. in real applications. IEEE Proceedings, 186-189.

BIERFERT, H., KIRSTEIN, M., LANCE D., REINHARD, C., RUPPRATH, RUPPRATH, R. (1981): Klassifikation phonetischer Signale. Institut für Kommunikationsforschung und Phonetik, Universität Bonn, Abschlußbericht zum Forschungsvorhaben T/RF 33/81031/81329.

BISIGNANI, WILLIAM T., BELL, JOHN O., MILLER, KENNETH H.: Speech intelligibility enhancement. Bericht MITRE Corporation, Metrek Division, Mclean, VA 22102, (Erscheinungsjahr unbekannt).

BLACK, J.W., LASHBROOK, W., NASH, E., OYER, H.J., PEDREY, C., TOSI, O.I., and TRUBY, H. (1973): Reply to Speaker Identification by speech spectrograms: Some further observations. Journal of the Acoustical Society of America, 54, 535-537.

BLAIN B.J. (1980): Transcription accuracy, British Home Office, PSDB, Technical Memorandum nr. 38/80.

BLAIN B.J., WALKER R.W., LEWIS S.R. (1980): Collection of speech samples. British Home Office, PSDB, Technical Memorandum Nr. 37/80.

BOGNER, Robert E. (1981): On talker verification via orthogonal parameters. In: IEEE-ASSP 29, 1-12.

BOLT, R.H., COOPER, F.S., DAVID, E.E., DENES, P.B., PICKETT, J. M., STEVENS, K.N. (1970): Speaker identification by speech spectrograms: A scientists`view of its reliability for legal purposes. Journal of the Acoustical Society of America, 47, 597-612.

BOLT, R.H.,COOPER, F.S., DAWID, E.E., DENES, P.B., PICKETT, J.M., STEVENS, K.N. (1973): Speaker Identification by speech spectrograms: some further observations. Journal of the Acoustical Society of America, 54, 531-534.

BRAHM, K. (1950): Die Analyse der s-Laute mit elektroakustischem Verfahren. Folia Phoniatrica, 2, 238-251.

BRICKER, P.D., GNANADESIKAN, R., MATHEWS, M.V., PRUZANSKY, S., TUKEY, P.A., WACHTER, K.W., WARNER, J.L. (1971): Statistical techniques for talker identification. Bell System Technical Journal, 50, 1427-1454.

BRICKER, P.D., PRUZANSKY, S. (1966): Effects of stimulus content and duration on talker identification. Journal of the Acoustical Society of America, 40, 1441-1449.

BRICKER, P.D., PRUZANSKY, S. (1976): Speaker Recognition. In: Lass, R. (Ed.): Contemporary Issues in Experimental Phonetics. Academic Press, New York/London, 295-326.

BRIDLE, J.S., PARTICIPANTS IN RSG 10, PROJECT ONE (1983): Connected Word Recognition for use in military systems. Final Report, Final Report of the First Project within a Research Programme in Speech Processing Conducted by NATO DRG AC 243/Panel III/RSG 10, Ref. AC/243(Panel III)D/218.

BRIDLE, J.S., BROWN, M.D., CHAMBERLAIN, R.M. (1981): An algorithm for connected word recognition, JSRU, Cheltenham, UK, Research Report Nr. 1010.

BRODERICK, P.K., PAUL, J.E., RENNICK, R.J. (1975): Semi-automatic speaker identification system. Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, Kentucky.

BUNDESKRIMINALAMT, Technische Forschung KI-2 (1979): Forschungsprojekt Forensische Sprechererkennung. Projektbericht Juni 1979, Wiesbaden.

BUNDESKRIMINALAMT, Technische Forschung KI-2 (1980): Forschungsprojekt Forensische Sprechererkennung. Projektbericht April 1980, Wiesbaden.

BUNDESKRIMINALAMT, Technische Forschung KI-2 (1981): Forschungsprojekt Forensische Sprechererkennung. Projektbericht März 1981, Wiesbaden.

CALINSKI, T., JASSEM, W., KACZMAREK Z. (1970): Investigation of vowel formant frequencies as personal voice characteristics by means of multivariate analysis of variance. In: Speech Analysis and Synthesis 2, 7-39, Warschau.

CARBONELL, J.R., GRIGNETTI, M.C., STEVENS, K.N., WILLIAMS, C.E., WOODS, B. (1965): Speaker authentication techniques. Report 1296 prepared under contract-nr. DA-28-043-AMC-00116(E) by Bolt, Beranek and Newman Inc., Cambridge, Mass.

CARROLL, J.D. (1972): Individual differences and multidimensional scaling. In R.N. Shepard, A.K. Romney, and S. Nerlove (Eds.), Multidimensional scaling: Theory and applications in the behavioral sciences. Vol.1: Theory. New York: Seminar Press, 105-155.

CHEUNG, R.S., EISENSTEIN, B.A. (1978): Feature selection via dynamic programming for text-independent speaker identification. IEEE - ASSP Z6: 397-403.

CHEUNG R.S. (1978): Feature Selection using adaptive learning networks for text independent speaker verification. Journal of the Acoustical Society of America, 68, 183.

CLARKE, F.R., BECKER, R.W. (1969): Comparison of techniques for discriminating among talkers. Journal of Speech and Hearing Research, 12, 747-761.

CLARKE, F.R., BECKER, R.W., NIXON, J.C. (1966): Characteristics that determine speaker recognition. Report ESD-TR-66-636, Electronic Systems Division. Air Force Systems Command, Hanscom Field.

CLARKE, F.R. R.W. (1969): Comparison of techniques for Discriminating among talkers. Journal of Speech and Hearing Research, 12, 747.

COHEN J.R., CRYSTA T.H., HOUSE A.S., NEUBURG E.P. (1980): Weighty voices and shaky evidence - a critique. Journal of the Acoustical Society of America, 1884-1886.

COLEMAN, R.O. (1971): Male and female voice quality and it's relationsship to vowel formant frequencies. Journal of Speech and Hearing Research, 14, 565-577.

COLEMAN, R.O. (1973): Speaker identification in the absence of intersubject differences in glottal source characteristics. Journal of the Acoustical Society of America, 53, 1741-1743.

COMPTON, A.J. (1963): Effects of filtering and vocal duration upon the identification of speakers, aurally. Journal of the Acoustical Society of America, 35, 1748-1752.

COOLEY, J.W., TUKEY, J.W. (1965): An Algorithm for the Machine Computation of Complex Fourier Series. Math. Comp. 19, April, 297-301.

DANILOFF, R.G., HAMMARBERG, R.E. (1973): On defining Coarticulation. Journal of Phonetics, I, 239-248.

DANTE, H.M., SARMA, V.V.S. (1979): Automatic Speaker Identification for a Large Population. IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, Nr. 3, Juni, 255-263.

DAS, S.K., Mohn, W.S. (1971): A scheme for speech processing in automatic speaker verification. I.E.E.E. Transactions on Audio and Electroacoustics, AU-19, 32-43.

DAVIS, R.L., HYDRICK, B.M., DODDINGTON, G.R. (1979): Total Voice Speaker Verification. Rome Air Development Center, Griffiss Air Force Base, New York 13441, Final Technical Report, RADC-TR-78-260.

DIEBOLD (1981): Management Report, Mai/Juni, 11-17.

DODDINGTON, G.R. (1970): A method of speaker verification. Paper presented at the 80th meeting of the Acoustical Society of America, November 3-8, Houston, Texas.

DODDINGTON, G.R., HYDRICK, B., BEEK, B. (1973): Some results on speaker verification using amplitude spectra. Paper presented at the 86th meeting of the Acoustical Society of America, October 30-November 2, Los Angeles, California.

DODDINGTON, G.R. (1976): Personal identity verification using voice. Paper on: Electro 76, Boston.

DODDINGTON, G.R., HYDRICK, B. (1976): "Total Voice"-Speaker Verification goes de luxe. Journal of the Acoustical Society of America, 96.

DODDINGTON, G.R., HYDRICK B. (1977): User adaptation in operational speaker verification. Journal of the Acoustical Society of America, 89.

DOHERTY, E.T. (1975): Evaluation of selected acoustic parameters for use in speaker identification. Journal of the Acoustical Society of America, 58, 107.

DUBES, R., JAIN A.K., TOSI, O.I. (1979): Speaker recognition studies on choral speech. Carnahan Conference on Crime

Countermeasures, University of Kentucky, Lexington, Kentucky, 165-171.

DUKIEWICZ, L. (1970): Frequency-band dependence of speaker identification. In: W. Jassem (Ed.), Speech analysis and synthesis, Vol. II, Warsaw, Poland: Institute of Fundamental Technical Research, Polish Academy of Sciences.

EL CHAFEI, C. (1980): A real time speaker recognition system on minicomputer. Carnahan Conference, University of Kentucky, Lexington, Kentucky, 53-56.

ENDRES, W., BAMBACH, W., FLOSSER, G. (1971): Voice spectrograms as a function of age, voice disguise, and voice imitation. Journal of the Acoustical Society of America, 49, 1842-1848.

FAEHRMANN, R., (1973): Die kriminalphonetische Expertise und der Sprechstimmensachverständige heute. Kriminalistik, 27.3, 97.

FANT, G. (1960): Acoustic Theory of Speech Production. Den Haag.

FANT, G. (1983): Phonetics and Speech Technology. Speech Transmission Laboratory, Quarterly Progress and Status Report, October 15, 20-35.

FEJFAR, A. (1977): Test Results - Advanced Development Models of BISS Identity Verification Equipment, Volume I, Executive Summary. MITRE Technical Report MTR-3442, 26 September.

FURUI, S., ITAKURA, F., Saito, S. (1972): Talker recognition by long-time averaged speech spectrum. Electronics and Communications in Japan, 55-A, 54-61.

FURUI, S. (1978): Effects of long-term spectral variability on speaker recognition. Journal of the Acoustical Society of America, 64, 183.

FURUI, S. (1979): New techniques for automatic speaker verification using telephone speech. Journal of the Acoustical Society of America, 66, 35.

FURUI, S. (1981): Cepstral analysis technique for automatic speaker verification. IEEE-ASSP 29, 254-272.

FURUI, S. (1981): Comparison of speaker recognition methods using statistical features. IEEE-ASSP 29, 342-350.

GARVIN, P.L., Ladefoged, P. (1963): Speaker identifikation and message identifikation in speech recognition. Phonetica 9, 193-199.

GEISER, G., SCHUMACHER, W. (1983): Sprachausgabe: Stand der Technik und ergonomische Gestaltung. Nachrichtentechnische Zeitschrift, Band 36, Heft 11, 740-744.

GEMELLI, A., PASTORI, G. (1934): L'analisi elettroacustica del linguaggio. Milano. (Zitiert nach TARNOCZY 1954).

GEPPERT, R., KUHN, M.H., PIOTROWSKY, H., TOMASCHEWSKI, H. (1979): An operational system for personnel authentication using long-term averaged voice spectrum. Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, Kentucky, 77-80.

GLAVE, R.D., VAN DER GIET, G. (1978): The DAWID Speech Recognition System. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Tulsa, 429-432.

GLAVE, R.D. (1977): Bewertungskriterien eines Spracherkennungssystems - am Beispiel des DAWID-Systems. Sprache und Datenverarbeitung, Nr. 2.

GLENN, J.W., KLEINER, N. (1968): Speaker identifikation based on nasal phonation. Journal of the Acoustical Society of America, 43, 368-372.

GOLD, B., RADER, C.M. (1969): Digital Processing of Signals. McGraw-Hill, New York, usw.

GOLDSTEIN, U.G. (1976): Speaker-identifying features based on formant tracks. Journal of the Acoustical Society of America, 59, 176-182.

GOYDKE, H. (1973): Untersuchungen zur nachträglichen Bestimmung des Frequenzgangeinflußes der Übertragungseinrichtungen auf Telefongespräche. Fortschritte der Akustik, Plenarvorträge und Kurzreferate der 3. Tagung der Deutschen Arbeitsgemeinschaft für Akustik, Aachen.

GRACE, H. Y.-K. (1981): Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. Journal of the Acoustical Society of America, 70, 966-975.

GRAY, C.H.G., KOPP, G.A. (1944): Voice print identifikation. Unpublished report, Bell Laboratories.

GRAY, A.H., MARKEL, J.D. (1976): Distance Measures for Speech Processing. IEEE ASSSP 24, Nr.5, 380-391.

GREEN, N. (1972): Automatic speaker recognition using pitch measurements in conversational speech. Joint Speech Research Unit, Report 1000, Cheltenham, England.

GREGORY, P.W., REAVES, J.M. (1980): Multistation voice data entry configuration study. Technology Service Corporation, Santa Monica, L.A. County, CA 90405, Report Nr. RADC-TR-81-50.

GRENIER, Y. (1980): Speaker adaption through canonical correlation analysis. IEEE International Conference on Acoustics, Speech and Signal Processing, Denver, April 9-11, 888-891.

HAIR, G.D., REKIETA, T.W. (1972): Techniques for objektive speaker identifikation. Paper presented at the Eighty-Fourth meeting of the Acoustical Society of America. November 28 - December 1, Miami Beach, Florida.

HALL, M., TOSI, O. (1975): Spectrographic and aural examination of professionally mimicked voices. Journal of the Acoustical Society of America, 58, 107.

HARRIS, S. (Editor) (1980): Voice-interactive systems: applications and payoffs. Proceedings of a symposium, organized by The Naval Air Development Center, 13-15, May.

HAWLEY, M.E. (Hrsg.) (1977): Speech intelligibility and speaker recognition, Stroudsbourg, Pennsylvania.

HEBERMAN, W., FEJFAR, A. (1976): Automatic identification of personnel through speaker and signature verification - system description and testing, Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, Kentucky, 23-30.

HARGREAVES, W.A., STARKWEATHER, J.A. (1963): Recognition of speaker identity. Languague and Speech, 6, 63-67.

HAZEN, B. (1973): Effects of differing phonetic contexts on spectrographic speaker identification. Journal of the Acoustical Society of America, 54, 650-660.

HECKER, M.H.L. (1971): Speaker recognition: An interpretive survey of the literature. ASHA Monographs, No. 16.

HEINZ, J.M., STEVENS, K.N. (1961): On the Properties of Voiceless Fricative Consonants. Journal of the Acoustical Society of America, 33, 589-596.

HELMS, R.E., DODDINGTON, G.R. (1976): Speaker verification

using passwords. Journal of the Acoustical Society of

HERMANN, L. (1894): Untersuchungen über die Konsonanten. Pflüger's Arch. ges. Physiol. 58, p. 255 u. p. 264 sowie 83 (1901) p. 1 u. p. 33. (Zitiert nach TARNOCZY 1954). America, 59, 96.

HOEFKER,U., JESORSKY,P., KRIENER, B., TALMI, M., Wesseling, D. (1979): A new system for authentication of voice. Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, Kentucky, 47-52.

HOLLIEN, H.(1974): Peculiar case of "voice prints" (L), Journal of the Acoustical Society of America, 56, 210-213.

HOLLIEN, H., Mc GLONE, R.E. (1976): An evaluation of the 'voice-print' techniques for speaker identification. Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, Kentucky.

HOLLIEN, H., MAJEWSKI, W. (1977): Speaker identification by long-term spectra under normal and distorted speech conditions. Journal of the Acoustical Society of America, 62, 975-980.

HOLMGREN, G.L. (1967): Physical correlates of speaker recognition. Journal of Speech and Hearing Research, 10, 57-66.

HUNT, M.J., YATES, J.W., BRIDLE, J.S. (1977): Automatic speaker recognition for use over communication channels.

HUNT, M.J. (1982): Work on Automatic Talker Recognition at JSRU, between 1972 and 1978, and comments on more recent work elsewhere. JSRU Research Report Nr. 1011.

HYDRICK, B., DODDINGTON, G.R. (1976): Speaker verification - an operational example. Journal of the Acoustical Society of America, 59, 96.

HYDRICK, B., DODDINGTON, G.R. (1976): Performance evaluation of speaker verification in entry control. Journal of the Acoustical Society of America, 64, 182.

INGEMANN, F. (1968): Identifikation of speaker's sex from voiceless fricatives. Journal of thee Acoustical Society of America 44, 1142-1144.

JESORSKY, P. (1976): Ein Vergleich verschiedener Parameterextraktionsverfahren der automatischen Spracherkennung.

5. IITB Kolloquium Mustererkennung, Februar.

JESORSKY, P. (1977): AUROS - Automatic recognition of speakers by computers feature-extraction and -reduction for text-dependent speaker-recognition. Proceedings of the 9 th International Congress of Acoustics, Madrid, Vol. I, Seite 102.

JOHNSON, C.C., DOHERTY, E.T. (1978): Speaker identification: new vectors for SAUSI. Journal of the Acoustical Society of America, 64, 182.

JOHNSON, C.C., HOLLIEN, H., DOHERTY, E.T. (1977): Long-term power spectra as speaker identification cue in simulated forensic situations. Journal of the Acoustical Society of America, 61, 70.

KASHYAP, R.L., (1976): Speaker recognition from an unknown utterance and speaker-speech interaction. IEEE-ASSP 24, 481-488.

KERSTA, L. (1962): Voiceprint identification. Nature 196, 1253-1257.

KIRSTEIN, Michael (1971): Akustische Untersuchungen zur automatischen Sprecheridentifikation. IPK-Forschungsberichte 34, Hamburg.

KOZHEVNIKOV, V.A., CHISTOVICH, L.A. (1965): Speech Articulation and Perception. Moskau-Leningrad, Englische Übersetzung: Clearinghouse for Federal Scientific and Technical Information, JPRS 30-543.

KRAUSE, H.J. (1976): Moeglichkeiten und Grenzen der Sprecheridentifizierung. Archiv füer Kriminologie, Band 157, 5. und 6. Heft, 154-164.

KUHN, M.H., GEPPERT, R. (1980): A low cost speaker identification device. Carnahan Conference on Crime Countemeasures, University of Kentucky, Lexington, Kentucky, 57-61.

LEA, W.A. (1982): Vortrag auf IEEE International Conference on Acoustics, Speech and Signal Processing, 3.-5. Mai, Paris.

LEA, W.A., SHOUP, J.E. (1979): Review of the ARPA SUR Project and survey of current technology in speech understanding. Final Report, Office of Naval Research, contract number N00014-77-C-0570, Speech Communications Research Laboratory, Los Angeles, CA. 90007.

LEE, Y.W. (1960): Statistical Theory of Communications.

John Wiley & Sons, New York, usw.

LEWIS, S.R. (1980): A technical introduction to the application of digital signal processing of speech for police use, British Home Office, PSDB, Technical Memorandum Nr. 32/80.

LI, K.P., DAMMANN, J.E., Chapman, W.D. (1966): Experimental studies in speaker verifikation, using and adaptive system. Journal of the Acoustical Society of America, 40, 966-978.

LI, K.P., HUGHES, G.W. (1972): Talker differences as they appear in correlation matrices from spectra of continuous speech. Paper presented at the Eighty-Fourth Meeting of the Acoustical Society of America, November 28-December 1, Miami Beach, Florida.

LI, K.P., KAU, C.K. (1987): Design of speaker identification schemes for large numbers of speakers. Journal of the Society of America, 63, 79.

LIM, J.S., OPPENHEIM, A.V. (1978): ALL-Pole Modeling of degraded Speech. IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, Nr. 3, June.

LIN, W.C., PILLAY, S.K. (1975): Selection of features and speech segments for speaker verification. Journal of the Acoustical Society of America, 58, 107.

LUCK, J.E. (1969): Automatic speaker verifikation using cepstral measurerements. Journal of the Acoustical Society of America 46, 1026-1032.

LUDEWIG, M. (1971): Vergleich von Sprechstimmen mittels der Sonagraphie. Dissertation, (HNO), München.

LUMMIS, R.C. (1973): Speaker verification by computer using speech intensity for temporal registration. I.E.E.E. Transactions on Audio and Electroacoustics. AU-21,80-89.

LUMMIS, R.C., Rosenberg, A.E. (1971): Test of automatic speaker verification method with intensively trained professional mimics. Paper presented at the Eighty-Second Meeting of the Acoustical Society of America. October 19-22, Denver, Colorado.

LUMMIS, R.C. (1972): Implementation of an on-line speaker verification scheme. Journal of the Acoustical Society of America, 52, 181.

MENZERATH, P., DE LACERDA, A. (1933): Koartikulation, Steuerung und Lautabgrenzung. Berlin, Bonn.

MARKEL, J.D., OSHIKA, B.T., GRAY, A.H. (1976): Long-term feature averaging in voice authentication. Journal of the Acoustical Society of America, 60, 13.

MARKEL, J.D., OSHIKA, B.T., Gray, A.H. (1977): Long term feature averaging for speaker recognition. IEEE-ASSP 25, 330-337.

MARKEL, J.D., DAVIS, S.B. (1979): Text-independent speaker recognition from a large linguistically unconstrained time-spaced data-base. IEEE-ASSP 27, 74-82.

MATHEWS, M.V., MILLER, J.E., David, E.E. (1961): Pitch synchronous analysis of voiced sounds. Journal of the Acoustical Society of America, 1961,33,179-186.

MATSUMOTO, H., HIKI, S., SONE, T., NIMURA, T. (1973): Multidimensional representation of personal quality and it's acoustical correlates. I.E.E.E. Transactions on Audio and Electroacoustics, AU-21,428-436.

McDERMOTT, B.J. (1964): Multidimensional analysis of circuit quality judgements. Journal of the Acoustical Society of America, 45, 774-781.

McGEHEE, F. (1937): The reliability of the identification of the human voice. The Journal of General Psychology, 17, 249-271.

McGEHEE, F. (1944): An experimental study of voice recognition. The Journal of General Psychology, 31, 53-65.

McGONEGAL, C.A., RABINER, L.R., ROSENBERG, A.E. (1979): The effects of ADCPM coding and LPC vocoding of an automatic speaker verification system. Journal of the Acoustical Society of America, 65, 131.

MEAD, K.O. (1974): Identification of speakers from fundamental frequency contours in conversational speech. Joint Speech Research Unit, Cheltenham, England, Report 1002.

MILLER, J.E. (1964): Decapitation and recapitation, a study of voice quality. Paper presented at the Sixty-Eighth Meeting of the Acoustical Society of America, October 21-24, Austin, Texas.

MOHN, W.S. (1971): Two statistical feature evaluation techniques applied to speaker identifikation. I.E.E.E. Transactions on

Computers, C-20, 979-987.

MOORE, R.K. (1977): Evaluating Speech Recognizers. IEEE ASSP 25, Nr. 2, 178-183.

MOYE, L.S. (1979): Study of the Effects on Speech Analysis of the Types of Degradation Occuring in Telephony. Standard Telecommunication Laboratories, Harlow, Essex, England.

MURRY, T., Cort, S. (1971): Aural identification of children's voices. Journal of Auditory Research, 11, 260-262.

MUTSCHLER, H., GEISER, G. (1983): Spracheingabe: Stand der Technik und ergonomische Gestaltung. Nachrichtentechnische Zeitschrift, Band 36, Heft 10, 672-677.

MUTSCHLER, H., (1981): Sprach-Ein- und - Ausgabe: Ergonomische Gestaltung für den praktischen Einsatz. Fraunhofer-Gesellschaft, Berichte, 1/2, 11-14.

NADC (1980): Voice-Interactive Systems: Applications and Payoffs. Conference Record, Editor LCDR Steve Harris.

NEIL, D.E., ANDREASON, T. (1981): Examination of voice recognition system to function in a bilingual mode. Naval Postgraduate School, Monterey, CA 93940, Report NPS55-81-004.

NEY, H. (1981): Bestimmung der Zeitverläufe von Intensität und Grundperiode der Sprache für die automatische Sprechererkennung. Frequenz 35, Heft 10, 265-270.

NGOC, C.B., MONBARON, J.J., MICHEL, J.G. (1983): An Integrated Voice Recognition System. IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-31, Nr.1, February.

NOLAN, F.J.D. (1980): The Phonetic Bases of Speaker Recognition. Dissertation, Fitzwilliam College, University of Cambridge.

NRC (National Research Council) (1979): On the Theory and Practice of Voice Identification. National Academy of Sciences, Washington D.C.

OCHIAI, Y. (1959): Phoneme and voice identification studies using japanese vowels. Language and Speech 2, 132-136.

OPPENHEIM, A.V., SCHAFER, R.W. (1975): Digital Signal Processing. Prentice-Hall, New Jersey.

OSGOOD, C.E., SUCI, G.J., TANENBAUM, P.H. (1957): The measurement of meaning. Urbana: Univ. of Illionis Press.

PAPCUN, G., LADEFOGED P. (1974): Two "voiceprint" cases.

Journal of the Acoustical Society of America, 55, 463.

PARSONS, T. W. (1978): Study and development of speech separation techniques. Rome Air Development Center, Griffiss Air Force Base, New York 13441, Final Technical Report, RADC-TR-78-105.

PAUL, J.F., RABINOWITZ, A.S., RIGANATI, J.P., RICHARDSON, J.M. (1975): Development of analytical methods for a semi-automatic speaker identification system. Carnahan Conference on Crime Countermeasures, University of Kentucky, Lexington, Kentucky.

PAUL, J,E. (1976): Semi automatic speaker identification system (SASIS). Final Report, Rockwell International, Anaheim, Cal.

PETERS, R.W. (1954): Studies in extra messages: Listener identification of speaker's voices under conditions of certain restrictions imposed upon the voice signal. U.S. Naval School of Aviation Medicine, Joint Project NM 001-064-01, Report 30, Pensacola, Florida.

PFEIFFER, L. (1974): Contributions of vowel and nasal sounds to speaker identification. Journal of the Acoustical Society of America, 55, 462.

POLLACK, I., PICKETT, J.M., SUMBY, W.H. (1954): On the identification of speakers by voice. Journal of the Acoustical Society of America, 26, 403-406.

POOCK, G.K. (1980): Experiments with Voice Input for Command and Control: Using Voice Input to Operate a Distributed Computer Network. Naval Postgraduate School, Monterey, CA 93940, Report Nr. NPS55-80-016.

POOCK, G.K., ARMSTRONG, J.W. (1981): Effect of Operator Mental Loading on Voice Recognition System Performance. Naval Postgraduate School, Monterey, CA 93940, Report Nr. NPS55-81-016.

POOCK, G.K. (1981): A Longitudinal Study of Computer Voice Recognition Performance and Vocabulary Size. Naval Postgraduate School, Monterey, CA 93940, Report Nr. NPS55-81-013.

PRUZANSKY, S. (1963): Pattern-matching procedure for automatic talker recognition. Journal of the Acoustical Society of America, 35,354-358.

PRUZANSKY, S. MATHEWS, M.V. (1964): Talker-recognition procedure based on analysis of variance. Journal of the Acoustical

Society of America, 36,2041-2047.

PTACEK, P.H., SANDER, E.K. (1966): Age recognition from voice. Journal of Speech and Hearing Research 9, 273-277.

PTACEK, P.H., SANDER, E.K., MALONEY, W.H., JACKSON, C.C. (1966): Phonatory and related changes with advanced age. Journal of Speech and Hearing Research, 9, 353-360.

RABINER, L.R., GOLD, B. (1975): Theory and Application of Digital Signal Processing, Prentice-Hall, New Jersey.

RABINER, L.R., SCHAFER, R.W. (1978): Digital Processing of Speech Signals. Prentice-Hall, New Jersey.

RAMISHVILI, G.S. (1966): Automatic voice recognition. Engineering Cybernetics 5, 84-90.

RATHBORN, H.A., BULL, H.R., CLIFFORD, B.R. (1981): Voice Recognition over the Telephone. Journal of Police Science and Administration, 9, Nr. 3, 280-284.

REICH, A.R. (1976): Effects of selected vocal disguises upon spectrographic speaker identification. Journal of the Acoustical Society of America, 59, 14.

ROSENBERG, A.E. (1973): Listener performance in speaker verification tasks. I.E.E.E. Transactions on Audio and Electroacoustics, AU-21,221-225.

ROSENBERG, A.E., SAMBUR, M. (1975): New techniques for automatic speaker verification. I.E.E.E. Transactions on Acoustics, Speech, and Signal Processing, ASSP-23,169-176.

ROSENBERG, A.E. (1976): Automatic Speaker Verification: A Review. Proceedings of the IEEE, 64, Nr. 4, April 1976, 475-487.

ROSENBERG, A.E. (1975): Evaluation of an automatic speaker verification system over telephone lines. Journal of the Acoustical Society of America, 57, 23.

ROSENBERG, A.E., SAMBUR M.R. (1974): An improved system for automatic speaker verification. Journal of the Acoustical Society of America, 55, 462.

ROSENBERG, A.E., SHIPLEY K.L. (1980): Speaker recognition in conjunction with speaker independent word recognition. 107.

ROTHMAN, H.B. (1975): Perceptual (aural) and spectrographic investigation of speaker homogeneity. Journal of the Acoustical Society of America, 58, 107.

RUSKE, G., SCHOTOLA, T. (1978): An approach to speech recognition using syllable decision units. IEEE Conference on Acoustics, Speech, and Signal Processing, Tulsa, Oklahoma, 722-725

RYAN, W.J., BURK, K.W. (1972): Predictors of age in the male voice. Paper presented at the Eighty-Fourth Meeting of the Acoustical Society of America, November 28 - December 1, Miami Beach, Florida.

SAKAI, T., NAKAGAWA, S. (1975): Continuous Speech Understanding System LITHAN. Studia Phonologica IX, 45-63.

SAKOE, H., CHIBA, S. (1978): Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE ASSP 26, Nr.1, 43-49.

SAMBUR, M.R. (1975): Selection of acoustic features for speaker identification. I.E.E.E. Transactions on Acoustics, Speech, and Signal Processing, ASSP-23, 176-182.

SAMBUR, M.R. (1976): Speaker recognition using orthogonal linear prediction, IEEE-ASSP 24, 283-289.

SARMA, V.S., VENUGOPAL, D. (1977): Performance evaluation of automatic speaker verification systems. IEEE-ASSP 25, 264-266.

SCHLITT, H., DITTRICH, F. (1972): Statistische Methoden der Regelungstechnik. BI-Hochschultaschenbücher, Mannheim, Wien, Zürich.

SCHWARZ, M., SHAW, L. (1975): Signal Processing. Discrete Spectral Analysis, Detection, and Estimation. McGraw-Hill, N.Y.

SCHWARTZ, M.F. (1968): Identification of speaker sex from isolated voiceless fricatives. Journal of the Acoustical Society of America, 43,1178-1179.

SCHWARTZ, M.F., RINE, H.E. (1968): Identification of speaker sex from isolated, whispered vowels. Journal of the Acoustical Society of America, 44, 1736-1737.

SECREST, B.G., HELMS, R.E. (1977): Remote terminal speaker verification. Rome Air Development Center, Griffiss Air Force Base, New York 13441, Final Technical Report, TR 08-77-07.

SHEARME, J.N., HOLMES, J.N. (1959): An experiment concerning the recognition of voices. Languague and Speech 2, 123-131.

SHIPP, T., HOLLIEN, H. (1969): Percepttion of the aging male voice. Journal of Speech and Hearing Research 12, 703-710.

SHRIDHAR M., VIDALON M. (1976): An algorithm for speaker verification. Journal of the Acoustical Society of America, 60, 13.

SMRKOWSKI, L. (1976): Study of speaker identification by aural and visual examination of non-contemporary speech samples. J. Assoc. official analytical chemists 59, 927-931.

SMRKOWSKI, L. (1977): Voice identification. Michigan department of state police.

SMITH, J.E.K. (1962): Decision-theoretic speaker recognizer. Paper presented at the Sixty-Fourth Meeting of the Acoustical Society of America, November 7-10, Seattle, Washington.

SMITH, C.P. (1979): Intelligibility threshold level (ITL) ratings of some current digital voice communications processors. Rome Air Development Center, Inhouse Report RADC-TR-79-215.

SOLI, Sigfrid D. (1981): Second formants in fricatives: Acoustic consequences of frikative-vowel coarticulation. Journal of the Acoustical Society of America, 70, 966-975.

SONDHI, M.M., SCHMIDT, C.E., RABINER, L.R. (1981): Improving the Quality of a Noisy Speech Signal. The Bell System Technical Journal, Vol. 60, Nr. 8, October, 1847-1857.

STEFFEN-BATOG M., JASSEM W., GRUSZKA-ROSCIELAK (1970): Statistical distribution of short-term F_0 values as a personal voice characteristic. Speech Analysis and Synthesis 2, 197-206, Warschau.

STERNBERG, S. (1969): Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 57,421-457.

STEVENS, K.N., WILLIAMS, C.E., CARBONELL, J.R., WOODS, B. (1968): Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material. Journal of the Acoustical Society of America, 44, 1596-1607.

STUMPF, C. (1926): Die Sprachlaute, Berlin, 1926.

SU, L.S., LI, K.P., Fu, K.S. (1974): Identification of speakers by use of nasal coarticulation. Journal of the Acoustical Society of America, 56, 1876-1882.

TARNOCZY, T. (1954): Die akustische Struktur der stimmlosen Engelaute. Acta Linguistica Hungaricae, 313-349

TARNOCZY, T. (1970): Die Sprechchormethode. Acustica, 23, Heft 4, 173-188.

TARNOCZY, T. (1971): Das durchschnittliche Energie-Spektrum der Sprache (für sechs Sprachen). Acustica, 24, Heft 2, 58-73.

TAYLOR, K.F. (1979): Investigators guide to voice identification and audio recording examination. Royal Canadian Mounted Police, Bericht.

TERANGO, L. (1966): Pitch and duration characteristics of the oral readings of males on a masculinity-femininity dimension. Journal of Speech and Hearing Research 9, 590-595.

TILLMANN, H.G. (1973): das individuelle Subjekt und seine persönliche Identität im phonetischen Kommunikationsprozeß. IKP-Forschungsberichte 48, Hamburg.

TOSI, O., OYER, H., LASHBROOK,W., PEDREY, C., NICOL, J. Nash, E. (1972): Experiment on voice identification. Journal of the Acoustical Society of America, 51, 2030-2043.

TOSI, O.I. (1979): Voice identification. Theory and legal applications. University Park Press, Baltimore.

TOSI, O.I. (1975): The problem of speaker identification and elimination, Sadanand Singh (Hrsg.): Measurement procedures in speech, hearing, and language. Baltimore, London, Tokyo, University Park Press, 399-431.

TRUBY, H.M. (1977): The first 100 years of "voiceprinting". Journal of the Acoustical Society of America, 62, 4.

TSCHESCHNER, W.(1964): Ergebnisse bei der Analyse von deutschen Sprachlauten. Zeitschrift für Phonetik, 17, Heft 1, 1-19.

UNGEHEUER, G. (1965): Ein einfaches Verfahren zur akustischen Verifikation von Sprechern. 5th International Congress on Acoustics, Paper A 17, Lüttich.

UNGEHEUER. G. (1966): Two-dimensional displays for speaker recognition. The Study of Sounds 12, 122-127.

VAN DER GIET, G. (1977): The application of the DAWID Speech Recognition System to voice programming and information querying. 9 th International Congress on Acoustics, Madrid 4.-9. Juli, Contributed Papers, Vol.1, 427.

VAN DER GIET, G. (1978): Anwendungen des DAWID-Spracherkennungssystems. Sprache und Datenverarbeitung, Nr. 2, 113-121.

VAN DER GIET, G. (1979): Dateneingabe in Informationssysteme mit dem DAWID Spracherkennungssystem. Forschungsbericht aus der

Wehrtechnik, BMVg-FBWT 79-22.

VAN DER GIET, G., BROSS, F., GRIGO, U., KUENZEL, H., SAUER, B., UNZICKER, R., Woestefeld, M. (1980): Rechnergestützte Sprechererkennung für Anwendungen im Polizeibereich. 5. Internationale Konferenz über spezielle Polizeitechnik, 233-252.

VAN DER GIET, G., Grigo, U., KUENZEL, H., SAUER, B., UNZICKER, R. (1980): Forensic speaker recognition system for police applications. International Conference: Security through Science and Engineering, Berlin, 50-57.

VAN DER GIET, G., KUENZEL H. J. (1981): Rechnergestützter Stimmenvergleich für forensische Anwendungen. Kriminalistik 9, 341-346.

VOIERS, W.D. (1964): Performance evaluation of speech processing devices II. The role of individual differences. Report AFCRL-66-24, Air Force Cambridge Research Laboratories, Office of Aerospace Research, Bedford, Massachusetts, December 1965. Proc. IEEE, 460-475.

VOIERS, W.D. (1964): Perceptual base of speaker identity. Journal of the Acoustical Society of America, 36, 1065-1073.

VOIERS, W.D., COHEN, M.F., MICKUNAS, J. (1965): Evaluation of speech processing devices I. Intelligibility, quality, speaker recognizability. Report AFCRL-65-826, Air Force Cambridge Research Laboratories, Office of Aerospace Research, Bedford, Massachusetts, July.

WAKITA, H. (1975): On the use of linear prediction error energy for speech and speaker recognition. Journal of the Acoustical Society of America, 57, 23.

WAKITA, H. (1976): Residual energy of linear prediction applied to vowel and speaker recognition. IEEE-ASSP 24, 270-271.

WASSON, D.A., DONALDSON, R.W. (1975): Speech amplitude and zero-crossings for automated identification of human speakers. IEEE-ASSP 23, 390-392.

WHITE G.M., NEELY, R.B. (1976): Speech-Recognition Experiments with linear Predication, Bandpass Filtering, and Dynamic Programming. IEEE-ASSP 24, Nr.2, 183-188.

WEINBERG, B., BENNET, S. (1971): A study of talker sex recog-

nition of esophageal voices. Journal of Speech and Hearing Research, 14, 391-395.

WEINBERG, B., BENNET, S. (1971): Speaker sex recognition of 5- and 6-year-old children's voices. Journal of the Acoustical Society of America, 50, 1210-1213.

WILLIAMS, C.E. (1964): The effects of selected factors on the aural identification of speakers. Sect.III of Report ESD-TDR-65-153, Electronics Systems Division, Air Force System Command, Hanscom Field.

WILSON, J.H. (1978): Application of the Fourier Series method to the detection and localization of signals embedded in a noise background. Journal of the Acoustical Society of America, 64(4), 1064-1069.

WOLF, J.J. (1971): Efficient acoustic parameters for speaker recognition. Journal of the Acoustical Society of America, 51, 2044-2055.

WOOD, C.A. (1978): Speaker identification by analysis of sound islands. Journal of the Acoustical Society of America, 64, 183.

WRENCH, E.H. (1980): Speaker authentication operational test and evaluation. Rome Air Development Center, Griffiss Air Force Base, New York 13441, (ITT Defense Communications Division), RADC-TR-80-64, April. Final Technical Report.

YOUNG, M.A., CAMPBELL, R.A. (1967): Effects of context on talker identification. Journal of the Acoustical Society of America, 42, 1250-1254.

An dieser Stelle möchte ich mich bei Herrn Prof. Dr. G. Ungeheuer für die Betreuung der Arbeit und sein Interesse bedanken. Durch seinen plötzlichen Tod konnte er ein Gutachten nicht mehr fertigen. Deshalb gilt mein besonderer Dank Herrn Prof. Dr. W. Lenders, der diese Aufgabe übernommen hat. Herrn Prof. Dr. H. P. Kelz danke ich für die Anfertigung des zweiten Gutachtens.

Herrn Prof. Dr. T. Mayer-Kuckuk und Herrn Prof. Dr. L. von Erichsen danke ich für die Bereitschaft zur Abnahme der Prüfungen in den Nebenfächern.

Bei der Amtsleitung des Bundeskriminalamtes, Herrn Dr. Boge und Herrn Boeden, bedanke ich mich für die Möglichkeit zur Durchführung dieser Arbeit.

Schließlich gilt mein besonderer Dank meinen Kollegen, Herrn Dr. H. J. Künzel und Herrn Dr. B. B. Borys (jetzt Universität Kassel), für zahlreiche Diskussionen und Anregungen.

Lebenslauf

Am 01. Februar 1947 wurde ich als Sohn der Eheleute Wilhelm und Gertrud van der Giet, geb. Voß, in Nieukerk, jetzt Kreis Kleve, geboren. Ich bin römisch-katholischer Konfession und besitze die deutsche Staatsbürgerschaft.

Nach dem Volksschulbesuch von Ostern 1953 bis Ostern 1957 besuchte ich das mathematisch-naturwissenschaftliche Gymnasium in Neukirchen-Vluyn und legte dort am 26. Februar 1966 die Reifeprüfung ab.

Im Sommersemester 1966 begann ich das Studium der Physik an der Universität Bonn. Dort legte ich am 25. Oktober 1968 die Diplomvorprüfung und am 14. Oktober 1971 die Diplomhauptprüfung für Physik ab. Die Diplomarbeit habe ich am Institut für Strahlen- und Kernphysik unter Leitung von Prof. Dr. T. Mayer-Kuckuk auf dem Gebiet der Kernreaktionen angefertigt.

Von Februar 1972 bis September 1978 war ich Mitarbeiter des Instituts für Kommunikationsforschung und Phonetik der Universität Bonn und arbeitete dort in einem Forschungsprojekt, das unter der wissenschaftlichen Leitung von Prof. Dr. G. Ungeheuer die Entwicklung eines Systems zur automatischen Erkennung gesprochener Sprache zum Ziel hatte. Im Rahmen dieser Arbeiten habe ich eine Dissertation auf dem Gebiet der artikulatorischen Phonetik angefertigt mit dem Titel "Ein neues Verfahren zur Messung von Artikulationsvorgängen". Am 25. September 1975 wurde mir der akademische Grad eines Doktors der Naturwissenschaften verliehen.

Die Aufgabenstellung des genannten Forschungsprojektes erforderte neben naturwissenschaftlichen auch tiefgehende Kenntnisse der Phonetik. Aus diesem Grund habe ich neben meinen Forschungsarbeiten zahlreiche Vorlesungen aus dem Gebiet der Kommunikationsforschung und Phonetik gehört und später auch ein entsprechendes

Studium durchgeführt. Während dieser Zeit hatte ich Gelegenheit, eigene Lehrveranstaltungen für Studierende dieses Faches durchzuführen.

Am 01. Oktober 1978 habe ich eine Tätigkeit am Bundeskriminalamt als Leiter eines Forschungsprojektes zum forensischen Stimmenvergleich aufgenommen. Im Rahmen der hier zu leistenden Forschungsarbeiten wurden die Experimente zu dieser Dissertation durchgeführt. Nach Beendigung der Förderung dieses Projektes bin ich seit dem 01. Juli 1983 in der Abteilung Datenverarbeitung des Bundeskriminalamtes im Bereich Kommunikation tätig.

Seit 1978 bin ich auf Veranlassung des Bundesministers der Verteidigung Mitglied in der Sprachverarbeitungsgruppe der NATO (DRG AC 243/Panel III/RSG-10), zuerst als Experte, später als deutscher Vertreter.

Am 28. November 1979 wurde ich vom Bundesminister des Innern zum Wissenschaftlichen Oberrat und am 22. Dezember 1980 zum Beamten auf Lebenszeit ernannt.